

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2001, 22:35:55 ; Search time 1157.88 Seconds
(without alignments)
4947.343 Million cell updates/sec

Title: US-09-656-668-198
Perfect score: 606
Sequence: 1 tgagttgcccttaccctccccc.....aagcctgtcttctgtcgcac 606
Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4736426750 residues
Word size : 0
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST.*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*
44: gb_est44:*
45: gb_est45:*
46: gb_est46:*
47: gb_est47:*
48: gb_est48:*
49: gb_est49:*
50: gb_est50:*
51: gb_est51:*
52: gb_est52:*
53: gb_est53:*
54: gb_est54:*
55: gb_est55:*
56: gb_est56:*
57: gb_est57:*
58: gb_est58:*
59: gb_est59:*
60: gb_est60:*
61: gb_est61:*
62: gb_est62:*
63: gb_est63:*
64: gb_est64:*
65: gb_est65:*
66: gb_est66:*
67: gb_est67:*
68: gb_est68:*
69: gb_est69:*
70: gb_est70:*
71: gb_est71:*
72: gb_est72:*
73: gb_est73:*
74: gb_est74:*
75: gb_est75:*
76: gb_est76:*
77: gb_est77:*
78: gb_est78:*
79: gb_est79:*
80: gb_est80:*
81: gb_est81:*
82: gb_est82:*
83: gb_est83:*
84: gb_est84:*
85: gb_est85:*
86: gb_est86:*
87: gb_est87:*
88: gb_est88:*
89: gb_est89:*
90: gb_est90:*
91: gb_est91:*
92: gb_est92:*
93: gb_est93:*
94: gb_est94:*
95: gb_est95:*
96: gb_est96:*
97: gb_est97:*
98: gb_est98:*
99: gb_est99:*
100: gb_est100:*
101: gb_est101:*
102: gb_est102:*
103: gb_est103:*
104: gb_est104:*
105: gb_est105:*
106: gb_est106:*
107: gb_est107:*
108: gb_est108:*
109: gb_est109:*
110: gb_est110:*
111: gb_est111:*
112: gb_est112:*
113: gb_est113:*
114: gb_est114:*
115: gb_est115:*
116: gb_est116:*

```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Wed Nov 7 09:21:19 2001

us-09-656-668-198.oli.rst

Page 4

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 405)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Kitzman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: www.bio.lnlnl.gov/dbp/image/image.html
Insert length: 500 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham.
Location/Qualifiers
FEATURES
source
1. 405
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1373409"
/clone_lib="NCI-CGAP_OV2"
/sex="female"
/tissue_type="ovary"
/lab_host="DH10B"
/note="Vector: pAMP10; mRNA made from invasive ovarian tumor, cDNA made by oligo-dr priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383."
BASE COUNT 118 a 102 c 79 g 106 t
ORIGIN

Query Match 45.4%; Score 275; DB 12; Length 405;
Best Local Similarity 99.7%; Pred. No. 4.5e-138; Indels 0; Gaps 0;
Matches 325; Conservative 0; Mismatches 1;
OY 272 ataaataaagacatgcccagtcctccctcaagtcctccgagccagggcctcaaggaat 331
Db 9 ATTAATAAAGACATCGCCAGTCCTCCCTCAAGCTCCGACGACGCTCAAGCCTCAAGCCT 68
OY 332 tccataacagtagaagaacactaaatattgattcaaaaatcctcaagcaactagaagaat 391
Db 69 TCCAAATTAAGAGTAGATGAACACTAATATGATTTCAAAATCTCAGCACTGAGAGAAAT 128
OY 392 gacaaacactcctgtgtgctgctgagctgtcctagtttagcatggaadtttcaagttc 451
Db 129 GACCAACATCTCTGTTGGCTCGGACTGCTAATGTTTGAAGTTTCAGTTTC 188
OY 452 caggaagacccctcaagcctgctgctgtgtcacccttagagagctgagggactctcaatc 511
Db 189 CAGGAAGCCCTCAGGCTCGGCTCTGCTGACCTTAGAGGCTGAGGAGCTTCATATAC 248
OY 512 agaattgctctgtgctgctgagtagaataacttaattgttaacatgtgaaacatc 571
Db 249 AGAATTAGTCTTGGCCACTGAGATGAATTAATTATTTGATACATGATGAAGAACTC 308
OY 572 tataaacactactgaagcctgtct 597
Db 309 TATAAACATCTACTGAAGCCTGTCT 334

RESULT 3
A1683094 473 bp mRNA EST 16-DEC-1999
LOCUS A1683094/c
DEFINITION tx01et12.x1 NCI-CGAP_Ut4 Homo sapiens cDNA clone IMAGE:2267950 3',
RNA sequence

ACCESSION A1683094
VERSION A1683094.1 GI:4893286
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 473)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: www.bio.lnlnl.gov/dbp/image/image.html
Insert length: 2912 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 405.
Location/Qualifiers
FEATURES
source
1. 473
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2267950"
/clone_lib="NCI-CGAP_Ut4"
/tissue_type="serous papillary carcinoma, high grade, 2 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr. Average insert size 1.48 kb. Life Technologies catalog #: 11542-016"
BASE COUNT 142 a 88 c 95 g 148 t
ORIGIN

Query Match 44.2%; Score 268; DB 23; Length 473;
Best Local Similarity 100.0%; Pred. No. 2.9e-134; Indels 0; Gaps 0;
Matches 268; Conservative 0; Mismatches 0;
OY 330 attcaataacagtagaataacactaaatattgattcaaaaatcctcaagcaactagaaga 389
Db 473 ATTCCAAATTAAGAGTAGATGAACACTAATATGATTTCAAAATCTCAGCACTGAGAGA 414
OY 390 atgacacacactcctggtgtgctgctgagctgtcctagtttagcatggaagttaggt 449
Db 413 ATGACCAACATCTCTGTTGGCTCGGCTGACCTGCTAATGTTTGAAGTTTCAGT 354
OY 450 tccaggaagcctcagcctggtgctgtgtccttagccttagcagctgagggactctcaat 509
Db 353 TCCAGGAAGCCTCAGGCTCGGCTGCTGCTGACCTAGCAGCTGAGGAGCTTCAAT 294
OY 510 acagatagctctgtgtgctgctgagtagaataacttaattgttaacatgtgaaacaa 569
Db 293 ACAGAAATAGCTTGTGTCACAGTAGATGAATTAATTATTTGAAACATGTAAGAAACA 234
OY 570 tctataaacactactgaagcctgtct 597
Db 233 TCTATAAACATCTACTGAGCCTGTCT 206

RESULT 4
A1307373 414 bp mRNA EST 08-APR-1999
LOCUS A1307373
DEFINITION tb26f01.x1 NCI-CGAP_K1d12 Homo sapiens cDNA clone IMAGE:2055461 3',
RNA sequence.
ACCESSION A1307373

VERSION A1307373.1 GI:4002098
EST.
KEYWORDS Homo sapiens
SOURCE human.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 414)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
Tumor Gene Index
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 592 Std Error: 0.00
Seq primer: -400P from Gibco.
Location/Qualifiers

FEATURES
source
1. .414
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2055481"
/clone_lib="NCI-CCAP-Kid12"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT73D-Pac (pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI-CCAP Kid5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneids 1323912-1325831, 1471368-1472903 and 1492104-1493255). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 126 a 80 c 85 g 123 t
ORIGIN

Query Match 33.7%; Score 204; DB 18; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.6e-99;
Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 394 ccaacacactggtgacctgagactgtctcattgacattgaagttcaagttca 453
|||||
DB 414 CCAACACACTGCTGGCTGGGACCTGCTCTAGTTTGAAGATTTCAGGTTCCA 355
|||||

OY 454 ggaagacctcagcctggtgctgtgtaccctcagcctgagagacttcaacag 513
|||||
DB 354 GGAAGACCTCAGCCTGGCTGCTGCTACCTTACAGCTGAGGACTCTCAATACAG 295
|||||

OY 514 aattagctcttgacacgagatgaatataacttaattgtaacatgtaaacactca 573
|||||
DB 294 AATTAGCTCTTGACACGAGATGAAATATATTATTTGTAACATGTAAGAAACATCTA 235
|||||

OY 574 taacacactcactgaagcctgtctc 597
|||||
DB 234 TAAACATCTACTGTAAGCCTGTTCT 211
|||||

RESULT 5
A1870547/c 398 bp mRNA EST 07-MAR-2000
LOCUS A1870547.w147a04.x1 NCI-CCAP_Ut1 Homo sapiens cDNA clone IMAGE:2428014.3.
DEFINITION mRNA sequence.
ACCESSION A1870547
VERSION A1870547.1 GI:5544515

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 398)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
Tumor Gene Index
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1826 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 373.
Location/Qualifiers

FEATURES
source
1. .398
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2428014"
/clone_lib="NCI-CCAP_Ut1"
/tissue_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"

BASE COUNT 130 a 56 c 68 g 144 t
ORIGIN

Query Match 12.7%; Score 77; DB 103; Length 398;
Best Local Similarity 100.0%; Pred. No. 1.4e-30;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 521 ctcttgacactgagatgataatacttaattgttaacatgtgaacacattacaat 580
|||||
DB 398 CTCTTGACACTGAGATGAAATACTTAAATTGTAACTGTGAACATCTATAACAT 339
|||||

OY 581 ctactgaagcctgtctc 597
|||||
DB 338 CTACTGAAGCCTGTCTC 322
|||||

RESULT 6
AM979007 269 bp mRNA EST 02-JUN-2000
LOCUS AM979007 MAGP resequences, MAGP Homo sapiens cDNA, mRNA sequence.
DEFINITION AM979007
ACCESSION AM979007
VERSION AM979007.1 GI:8170290
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 269)
Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and Quackenbush, J.
Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray
Unpublished (2000)
CONTACT: John Quackenbush
The Institute for Genomic Research

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 530)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLW at:
www-bio.lnl.gov/bbrp/image/image.html
Insert Length: 576 Std Error: 0.00
Seq primer: -400P from Gtlco
High quality sequence stop: 404
POLYA-No.

FEATURES
SOURCE 1.530
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2261007"
/clone_lib="NCI-CGAP_OV35"
/tissue_type="tumor, 5 pooled (see description)"
/lab_host="DH10B"
/note="Organ: ovary; Vector: PCMV-SPORT6; Site:1; Salt:
Site:2: NotI; This library represents the normalized
version of NCI-CGAP_OV23. Cloned unidirectionally.
Primer: Oligo dT. Average insert size 0.86 kb. Tumor
types include: mixed Mullerian tumor, papillary serous,
clear cell, spindle cell. All are primary tumors,
metastasis positive. Constructed by Life Technologies."

BASE COUNT 131 a 137 c 63 g 198 t 1 others
ORIGIN

Query Match 4.1%; Score 25; DB 22; Length 530;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 gtgggaacatgctactgcatccta 99
|||||
DB 105 GTGGGAACATGCTACTGCACTCA 81

RESULT 10
LOCUS AW771618 541 bp mRNA EST 04-MAY-2000
DEFINITION hn59c06.x1 NCI-CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3032170 3',
mRNA sequence.
ACCESSION AW771618
VERSION AW771618.1 GI:7703676
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 541)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.

FEATURES
SOURCE 1.541
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3032170"
/clone_lib="NCI-CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site:1: Not I; Site:2: Eco RI;
Plasmid DNA from the normalized library NCI-CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneids 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."
cdna Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLW, send email to:
info@image.lnl.gov
Seq primer: -400P from Gtlco
High quality sequence stop: 463.

BASE COUNT 136 a 136 c 65 g 204 t
ORIGIN

Query Match 4.1%; Score 25; DB 120; Length 541;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 gtgggaacatgctactgcatccta 99
|||||
DB 101 GTGGGAACATGCTACTGCACTCA 77

RESULT 11
LOCUS BE144594 323 bp mRNA EST 21-JUN-2000
DEFINITION MR0-HT0167-141199-002-C02 HT0167 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE144594
VERSION BE144594.1 GI:8607227
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 323)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Brines,M.R.,
Nagel,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalhal,A.F., Matsukuma,A., Bala,G.S., Simpson,D.B.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
JOURNAL MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/ILICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l=612-MR0-HT0167-141>
199-002-c02&t3=1999-11-14&t4=1)
Seq primer: puc 18 forward

High quality sequence start: 13
High quality sequence stop: 323.
Location/Qualifiers

FEATURES
SOURCE

1. 323
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0167"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
low stringency conditions" 1 others
103 a 75 c 79 g 65 t

BASE COUNT
ORIGIN

Query Match 3.8%; Score 23; DB 164; Length 323;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 163 agaattactagccccaatgctc 185
|||||
db 135 AGAATTATCTAGCCCAATGCTC 157

RESULT 12 390 bp DNA GSS 13-JUL-1996
B03713
LOCUS
DEFINITION
CSRL-185e7-u CSRL flow sorted Chromosome 11 specific cosmid Homo
sapiens genomic clone CSRL-185e7, DNA sequence.

ACCESSION
B03713
KEYWORDS
GSS.
human.
ORGANISM
Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS
1 (bases 1 to 390)
Evans, G.A., Burbee, D., Davies, C., Garner, L., Oliver, T., Gilbert, M.,
Jones, D., Ward, T., Gillilan, E., Schagemann, J., Probst, S., Harris
, J., Depord, J., McFarland, J., Burzinski, K., Khan, M., Kupfer, K. and
Garner, H.R.

TITLE
JOURNAL
COMMENT
Genomic Sequence Sampled Map of Chromosome 11
Unpublished (1996)
Contact: Evans GA, Shane Probst
Mcdermott Center for Human Growth and Development
University of Texas Southwestern Medical Center At Dallas
5323 Harry Hines Blvd, Dallas TX 75235-8591
Tel: 214-648-1600
Fax: 214-648-1666
Email: g.evans@uts.wmed.edu, shanemcdermott.wmed.edu

PCR Primers
FORWARD: ACAAGAATTATCTAGCCCC
BACKWARD: AACCAACAACCTCACACAC
Seq primer: 77
Class: cosmid ends
High quality sequence stop: 390.
Location/Qualifiers

FEATURES
SOURCE

1. 390
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CSRL-185e7"
/clone_lib="CSRL flow sorted Chromosome 11 specific
cosmid"
/sex="female"
/cell_type="chimeric hamster somatic cell hybrid"
/note="Vector: scos-1; Human Chromosome 11 specific cosmid
library prepared from flow sorted human Chromosome 11
derived from Chinese Hamster Ovary (CHO) monochromosomal
somatic cell hybrid, j1"

BASE COUNT 85 a 111 c 70 g 113 t 11 others
ORIGIN

Query Match 3.8%; Score 23; DB 256; Length 390;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 163 agaattactagccccaatgctc 185
|||||
db 51 AGAATTATCTAGCCCAATGCTC 73

RESULT 13 528 bp DNA GSS 27-APR-1999
A0282338/c
LOCUS
DEFINITION
RPC111-79011.TV RPC1-11 Homo sapiens genomic clone RPC1-11-79011,
DNA sequence.

ACCESSION
A0282338
KEYWORDS
GSS.
human.
ORGANISM
Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS
1 (bases 1 to 528)
Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and Ventier, J.C.
Use of human BAC End Sequences for Sequence-Ready Map Building
Unpublished (1998)
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamas@tigr.org

Library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: 47
Class: BAC ends.
Location/Qualifiers

FEATURES
SOURCE

1. 528
/organism="Homo sapiens"
/db_xref="GDB:7530298"
/db_xref="taxon:9606"
/clone="RPC1-11-79011"
/clone_lib="RPC1-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC library"

BASE COUNT
ORIGIN

162 a 96 c 135 g 135 t
Query Match 3.8%; Score 23; DB 226; Length 528;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 163 agaattactagccccaatgctc 185
|||||
db 500 AGAATTATCTAGCCCAATGCTC 478

RESULT 14 422 bp mRNA EST 06-JUN-1995
R77087
LOCUS
DEFINITION y164904.r1 Soares placenta ND2HP Homo sapiens cDNA clone

Wed Nov 7 09:21:19 2001

us-09-656-668-198.oli.rst

IMAGE:144054 5' similar to contains LTR4 repetitive element ;, mRNA sequence.

ACCESSION R77087
 VERSION R77087.1 GI:851719
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 422)
 Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M., Holman M., Hultman M., Kucaba T., Le M., Lennon G., Merra M., Parsons J., Rikkin L., Rohlfing T., Soares M., Tan F., Trevisakis E., Waterston R., Williamson A., Wohlmann P. and Wilson R.
 The Mashu-Merck EST Project
 Unpublished (1995)
 CONTACT: Milson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 587
 High quality sequence stops: 333
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
 Putative full length read
 Insert length: 587 Std Error: 0.00
 Seq primer: M13Kp1
 High quality sequence stop: 333.

FEATURES
 Location/Qualifiers
 1..422
 /organism="Homo sapiens"
 /db_xref="GDB:553380"
 /db_xref="taxon:9606"
 /clone="IMAGE:144054"
 /clone_id="Soares placenta ND2HP"
 /sex="Female"
 /dev_stage="placenta obtained at birth (full term)"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: placenta; Vector: p1773D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTGAGAGATTCGGCGGCGAGATTTTCTTTTCTT 3'], (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p1773 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 102 a 93 c 88 g 134 t 5 others

ORIGIN

Query Match 3.5%; Score 21; DB 188; Length 422;
 Best Local Similarity 100.0%; Pred. No. 3.5;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 245 tgcccttgctgctgctctgt 265
 ||||||||||||||||||
 Db 350 tgccttgcctgcctgcctct 370

RESULT 15
 AL513616 1018 bp mRNA EST 13-FEB-2001
 LOCUS AL513616 LTI_NFL006_PL2 Homo sapiens cDNA clone XCL08B0012E07 5
 DEFINITION prime, mRNA sequence.
 ACCESSION AL513616
 VERSION AL513616.1 GI:12777110
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1018)
 Li W.B., Gruber C., Jesse J. and Polayes D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 CONTACT: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
 Location/Qualifiers
 1..1018
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="XCL08B0012E07"
 /clone_id="LTI_NFL006_PL2"
 /issue_type="placenta"
 /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end was enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Peng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : filiang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 212 a 308 c 318 g 176 t 4 others

ORIGIN

Query Match 3.5%; Score 21; DB 105; Length 1018;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 459 gccctcagcctgagctgctg 479
 ||||||||||||||||||
 Db 31 gccctcagcctgagcctgctg 11

Search completed: November 5, 2001, 23:19:10
 Job time: 2595 sec

Wed Nov 7 09:21:19 2001

us-09-656-668-198.oli.rst

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

M nucleic - nucleic search, using sw model

run on: November 5, 2001, 22:16:10 ; Search time 1182.83 Seconds
(without alignments)
7924.601 Million cell updates/sec

title: US-09-656-668-198
erfect score: 606
sequence: 1 tgagttgcccccttaccac.....aagcctgttctgtctgcac 606

coring table: IDENTITY_NUC
Gapop 10.0 / Gapext 1.0

searched: 1344157 seqs, 7733874588 residues
total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenBml:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_iny:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_hum1:*
31: em_htg_hum2:*
32: em_htg_rod:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v1:*
59: gb_v12:*
60: gb_v12:*
61: gb_v12:*
62: gb_v12:*
63: gb_v12:*
64: gb_v12:*
65: gb_v12:*
66: gb_v12:*
67: gb_v12:*
68: gb_v12:*
69: gb_v12:*
70: gb_v12:*
71: gb_v12:*
72: gb_v12:*
73: gb_v12:*
74: gb_v12:*
75: gb_v12:*
76: gb_v12:*
77: gb_v12:*
78: gb_v12:*
79: gb_v12:*
80: gb_v12:*
81: gb_v12:*
82: gb_v12:*
83: gb_v12:*
84: gb_v12:*
85: gb_v12:*
86: gb_v12:*
87: gb_v12:*
88: gb_v12:*
89: gb_v12:*
90: gb_v12:*
91: gb_v12:*
92: gb_v12:*
93: gb_v12:*
94: gb_v12:*
95: gb_v12:*
96: gb_v12:*
97: gb_v12:*
98: gb_v12:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	606	100.0	606	10	AX093380 Sequence
2	104.6	17.3	144014	66	AC021763 Homo sapi
3	104.6	17.3	159150	83	AP002438 Homo sapi
4	104.6	17.3	187200	82	AP001379 Homo sapi
5	97.2	16.0	152081	86	AC007381 Homo sapi
6	95.6	15.8	93714	79	AL161661 Homo sapi
7	95.6	15.8	147815	92	HS99886 Human DNA
8	93.6	15.4	185143	69	AC025861 Homo sapi

OY		61	atgttcaacctgggtgaggaaacatgctactcgtgatctlaatgcataagagggaagttaatgc	120
Db		61	ATTGTCAACCTGGGCGTGGGAACATCTACTGGCATCTAATGCATGTAGAGGCGATTAATGC	120
OY		121	tgtcaaatctctttcaacgcacagacagaagcccccacaaaagsaatltatcagccccaa	180
OY		121	TGCTAAACATCTTTTCACGCCACAGCAGAACAGCCCCCACAAAAGAATTAATCTTAGCCCCA	180
Db		121	TGCTAAACATCTTTTCACGCCACAGCAGAACAGCCCCCACAAAAGAATTAATCTTAGCCCCA	240
OY		181	atcttcataacactcgtcgttgtagaaaacctaccagagaaatcttactcgtgcttacaagta	240
Db		181	ATGTCCATAACCTGCTGTGAGAAAACTPACCGCAGATCTTACTGGGCTCATAGGTA	240
OY		241	agcttgccttgtctcgtgcttcctgtagataataaataaagacactgcagctccctcc	300
Db		241	AGCTTGCTTGTCTTGGGCTTCTGAAATATATAAAATAAGACACATGCCACCTCCCTCC	300
OY		301	cctcaagctcccgagcgagggctccaaggaattccaatacagtagaatgaagaacataa	360
Db		301	CCTCAACGTCCTCGAGCCAGCGGCTCAAGGCAATTCCTCAATACATAAATGAACACTTAATA	360
OY		361	tgtgatcttaaatctcagcaactagaaagaatggccacacactcctgtgtgctcgtggaatg	420
Db		361	TTGATTTTCAAATTTOTGCACTAGAAATAAGCAACAACCTCTGGTTGGCTGTGGGACTG	420
OY		421	tccatgatttagcatgtaaaaatttcaggcttcacagaagccctcacagcctgtgctgtg	480
Db		421	TCTTACTTTTAGCATTTGAAAGTTTCAGGTTCCAGGAAGCCCTCAGGCGCTGGGCTGCTG	480
OY		481	tcaacctagcagcgtagaggaactcttcaatacagaattagtctcttjggaactcggagatgaa	540
Db		481	TCACCCCTACAGACTGAGGACTCTTCOAATACAGAAATGAGCTTTGGCACTGGAGATGAA	540
OY		541	tataacttaacttgtaaacatgtgnaaacatctataacaactctaataagcctgttctt	600
Db		541	TATACCTTAATTGTATACATGTGAAAAACATATATAACATCTACTAGAACCTGTTCTGT	600
OY		601	ctgcac	606
Db		601	CTGCAC	606

AC0211763/C	LOCUS	AC0211763	144014 bp	DNA	HTG	13-OCT-2000
DEFINITION	Sequence, 17 unordered pieces.	AC0211763				WORKING DRAFT
ACCESSION	AC0211763.3	GI:10800280				
VERSION	HTG; HTGS_PHASE1; HTGS_DRAFT.					
KEYWORDS	human.					
SOURCE	homo sapiens					
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
REFERENCE	1 (bases 1 to 144014)					
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.					
TITLE	Homo sapiens chromosome 18, clone RP11-56021					
JOURNAL	Unpublished					
REFERENCE	2 (bases 1 to 144014)					
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,B., Batta,N., Beckert,R., Beda,F., Bonafant,L., Boukhalter,B., Brown,A., Burkett,G., Castle,A., Choepeil,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeBellano,K., Dewar,K., Domino,M., Doyle,M., Fensholt,J., Ferrelita,P., Fitzhugh,N., Forrest,C., Gage,D., Galagan,J., Gargano,S., Grant,G., Hagos,B., Heatford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kamm,L., Karatsis,A., Klein,J.,					
	Howland,J.C., Johnson,R., Levine,R., Lien,C., Liu,G., Locke,K.,					

13-Oct-2000
HTG
18, WORKING DRAFT

CG021763 144014 bp DNA
Homosapiens chromosome 18 clone RP11-56021 map 18, WORKING DRAFT
SEQUENCE, 17 unordered pieces.
CG021763
CG021763.3 GI:10800280
HTG; HTGS_PHASE1; HTGS_DRAFT.

Human.
Homosapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 144014)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homosapiens chromosome 18, clone Rpl1-56021
2 (bases 1 to 144014)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckert,J., Biedt,F.,
Boguslavskiy,L., Bouhagbaler,B., Brown,A., Burrett,G., Castle,A.,
Chopel,I., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeRella,P., Fitzhugh,K., Dewar,K., Domini,M., Doyle,M., Fennestor,J.,
Ferreira,P., Fitzhugh,N., Forrest,C., Gage,D., Galagan,J.,
Gardaya,S., Grant,G., Hagos,B., Jones,C., Kann,L., Karatas,A., Klein,J.,
Howland,J.C., Johnson,R., Jones,C., Levine,R., McGurk,A., McKernan,K.,
Lander,E., Lehoczeky,J., Levine,R., Mcewan,P., McGurk,A., McKernan,K.,
Macdonald,P., Marquis,N., Mennis,L., Morrow,J., Naylor,J.,
McPheeters,R., Weidim,J., O'Donnell,P., Olivari,T.M., Peterson,K.,
Norman,C.H., O'Connor,T., O'Donnell,P., Raymond,C., Riley,R., Rothman,D.,
Pierre,N., Pisanil,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Strange-Thomann,N.,

us-09-656-668-198.rge

Wed Nov 7 09:21:19 2001

VERSTON KEYWORDS SOURCE ORGANISM

AP002438.1 GI:8307742
HTG: HTGS-PHASE1; HTGS-DRAFT.
Homo sapiens DNA, clone: RP11-679N11.

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 159150)
Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
Homo sapiens 159,150 genomic DNA of 18q21
Published Only in DataBase (2000) In press

TITLE JOURNAL REFERENCE AUTHORS

2 (bases 1 to 159150)
Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (02-JUN-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
Kitasato Univ., 1-15-1 Kitasato, Sagami-hara, Kanagawa 228-8555,
Japan (E-mail: hattori@gsc.riken.go.jp,
URL: http://hgp.gsc.riken.go.jp/, Tel: 81-42-778-9923,
Fax: 81-42-778-9924)

COMMENT

Center: RIKEN Genomic Sciences Center (GSC)

Center code: RIKEN

Web site: <http://hgp.gsc.riken.go.jp/>

Contact: hattori@gsc.riken.go.jp

----- Project Information

Center project name: Humdraft18

Center clone name: RP11-679N11

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 141170 bases at least Q40

Consensus quality: 150107 bases at least Q30

Consensus quality: 153808 bases at least Q20

Insert size: 155950; sum-of-contigs

Quality coverage: 4.52x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 33 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 18093 contig of 18093 bp in length
18194 34028 contig of 15835 bp in length
34129 47853 contig of 13725 bp in length
47954 61375 contig of 13422 bp in length
61476 59763 contig of 8288 bp in length
69864 77424 contig of 7561 bp in length
77525 83516 contig of 5992 bp in length
83617 89288 contig of 5672 bp in length
89389 94381 contig of 4993 bp in length
94482 100220 contig of 5739 bp in length
100321 105057 contig of 4737 bp in length
105158 105157 contig of 100 bp
105158 110566 contig of 5408 bp in length
110566 110566 contig of 100 bp
110566 115962 contig of 5297 bp in length
115963 116062 contig of 100 bp
116063 119619 contig of 3557 bp in length
119620 119719 contig of 100 bp
119720 124300 contig of 4581 bp in length
124301 124400 contig of 100 bp
124401 127169 contig of 2769 bp in length
127170 127269 contig of 100 bp
127270 129941 contig of 2672 bp in length
129942 130041 contig of 100 bp
130042 132701 contig of 2660 bp in length
132702 132801 contig of 100 bp
132802 135223 contig of 2422 bp in length
135224 135323 contig of 100 bp
135324 138940 contig of 3617 bp in length
138941 139040 contig of 100 bp
139041 140352 contig of 1312 bp in length
140353 140452 contig of 100 bp
140453 142867 contig of 2415 bp in length
142868 142967 contig of 100 bp
142968 144444 contig of 1477 bp in length
144445 144544 gap of 100 bp
144545 146121 contig of 1577 bp in length
146122 146221 contig of 100 bp
146222 147660 contig of 1439 bp in length
147661 147760 gap of 100 bp
147761 149057 contig of 1297 bp in length
149058 149157 gap of 100 bp
149158 150648 contig of 1491 bp in length
150649 150748 gap of 100 bp
150749 152399 contig of 1651 bp in length
152400 152499 contig of 100 bp
152500 154103 contig of 1604 bp in length
154104 154203 gap of 100 bp
154204 155806 contig of 1603 bp in length

```

us-09-656-668-198.rge

Wed Nov 7 09:21:19 2001

* 155807 155906: gap of 100 bp
 * 155907 157204: contig of 1298 bp in length
 * 157205 157304: gap of 100 bp
 * 157305 158735: contig of 1431 bp in length
 * 158736 158835: gap of 100 bp
 * 158836 159150: contig of 315 bp in length.

FEATURES

source
 1. .159150
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18q21"
 /clone="RP11-679N11"
 misc_feature
 1. .18093
 /note="assembly_fragment"
 misc_feature
 18194. .34028
 /note="assembly_fragment"
 misc_feature
 34129. .47853
 /note="assembly_fragment"
 misc_feature
 47954. .61375
 /note="assembly_fragment"
 misc_feature
 61476. .69763
 /note="assembly_fragment"
 misc_feature
 69864. .77424
 /note="assembly_fragment"
 misc_feature
 77525. .83516
 /note="assembly_fragment clone_end:sp6 vector_side:left"
 misc_feature
 83617. .89286
 /note="assembly_fragment"
 misc_feature
 89389. .94381
 /note="assembly_fragment"
 misc_feature
 94482. .100220
 /note="assembly_fragment"
 misc_feature
 100321. .105057
 /note="assembly_fragment"
 misc_feature
 105158. .110565
 /note="assembly_fragment"
 misc_feature
 110666. .115962
 /note="assembly_fragment"
 misc_feature
 116063. .119619
 /note="assembly_fragment"
 misc_feature
 119720. .124300
 /note="assembly_fragment"
 misc_feature
 124401. .127169
 /note="assembly_fragment"
 misc_feature
 127270. .129941
 /note="assembly_fragment"
 misc_feature
 130042. .132701
 /note="assembly_fragment"
 misc_feature
 132802. .135223

Query Match 17.3%; Score 104.6; DB 83; Length 159150;
 Best Local Similarity 64.9%; Pred. No. 6.7e-20;
 Matches 155; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
 QY 24 cccagtaatttcgaattcctaagacgctgttttgattgtcacacctgggtggggaac 83
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 24951 CCCAGTAAATATTTCGCAATCCCGAGGACATTTTGGTGTGCACACGGGAACAGGAAC 25010
 QY 84 atgctactggcctcattgacatagagggcagtaagtctgctaaacatcttcaacgcaca 143
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 25011 CTGCTACTGGCATCTAATAGTAGAGGTGAAGGATGCTGCTGAACATCTTACATGTCACA 25070
 QY 144 ggcagagcccccacacagagaataattatctagcccccaaaatgtccatacaactgtctgtgag 203
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 25071 GGTGAGACCCCAATACAAAGAAATGATCCAGCCCAAAATATCATTAGTCCGAGGCTGAG 25130
 QY 204 aaacacctaccgcaggatcttactgggttcattaggttaagcttgcctttgttctggttc 262
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 25131 AAACCTGTTCTAAGGCGCATTTGCAAAAGTACCAATGAGAGTGTGCTGTGAAAGCTGGTTTC 25189

RESULT 4

AP001379 187200 bp DNA HTG 15-JUL-2000
 Homo sapiens chromosome 18 clone RP11-850H3 map 18q12, WORKING
 DRAFT SEQUENCE, 21 unordered pieces.
 AP001379 3 GI:9229958
 AP001379.3 HTGS_PHASE1; HTGS_DRAFT.
 HTG; HTGS_PHASE1; HTGS_DRAFT.
 Homo sapiens DNA, clone:RP11-850H3.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 187200)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 Homo sapiens 187,200 genomic DNA of 18q12
 Published Only in DataBase (2000) In press
 2 (bases 1 to 187200)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 Direct Submission
 Submitted (09-MAR-2000) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC):
 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
 (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/
 Tel:81-42-778-9923, Fax:81-42-778-9924)
 On Jul 15, 2000 this sequence version replaced gi:8117292.
 ----- Genome Center
 Center: RIKEN Genomic Sciences Center(GSC)
 Center code: RIKEN
 Web site: http://hgp.gsc.riken.go.jp/
 Contact: hattori@gsc.riken.go.jp
 ----- Project Information
 Center project name: HumDraft18
 Center clone name: RP11-850H3
 ----- Summary Statistics
 Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 179971 bases at least Q40
 Consensus quality: 182832 bases at least Q30
 Consensus quality: 184194 bases at least Q20
 Insert size: 185200; sum-of-contigs
 Quality coverage: 9.46x in Q20 bases; sum-of-contigs

 NOTE: This is a 'working draft' sequence. It currently consists of 21 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved
 1 33581 contig of 33581 bp in length
 33682 58093 contig of 24412 bp in length
 58194 77625 contig of 19432 bp in length
 77726 97031 contig of 19306 bp in length
 97132 111990 contig of 14859 bp in length
 112091 125465 contig of 13375 bp in length
 125566 135454 contig of 9889 bp in length
 135555 144628 contig of 9074 bp in length
 144729 151052 contig of 6334 bp in length
 151163 155511 contig of 4349 bp in length
 155612 160186 contig of 4575 bp in length
 160287 164088 contig of 3802 bp in length
 164189 167739 contig of 3551 bp in length
 167840 171188 contig of 3349 bp in length
 171289 174981 contig of 3693 bp in length
 175082 177828 contig of 2747 bp in length
 177929 180660 contig of 2732 bp in length
 180761 183192 contig of 2432 bp in length
 183293 185028 contig of 1736 bp in length
 185129 185875 contig of 747 bp in length
 185976 187200 contig of 1225 bp in length.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

[illegible]FEATURES
SOURCE

```
1. .18/200
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q12"
/clone="p011-850H3"
```

misc_feature	1.35591	/note="assembly_fragment"
misc_feature	33682.58093	/note="assembly_fragment"
misc_feature	58194.77625	/note="assembly_fragment"
misc_feature	77726.97031	/note="assembly_fragment"
misc_feature	97132.111990	/note="assembly_fragment"
misc_feature	112091.125465	/note="assembly_fragment"
misc_feature	125566.135454	/note="assembly_fragment"
misc_feature	135555.144628	/note="assembly_fragment"
misc_feature	144729.151062	/note="assembly_fragment"

```

151163. .155511
/note="assembly_fragment"
155612. .160186
/note="assembly_fragment"
160287. .164088
/note="assembly_fragment"
164189. .167739
/note="assembly_fragment"
167840. .171188
/note="assembly_fragment"
171289. .174981
/note="assembly_fragment"
175082. .177828
/note="assembly_fragment"
177929. .180660
/note="assembly_fragment"
180761. .183192
/note="assembly_fragment"
183293. .185028
/note="assembly_fragment"
185129. .185875
/note="assembly_fragment clone_end:T7 vector_side:right"
185976. .187200
/note="assembly_fragment"
39996 g 41403 g 52571 t 2000 others
BASE COUNT 51230 a
ORIGIN

```

		Query Match	17.3%;	Score 104.6;	DB 82;	Length 187200;			
		Best Local Similarity	64.9%;	Pred. No. 6.8e-20;		Gaps	0;		
		Matches 155;	Conservative	0;	Mismatches 84;	Indels	0;		
Qy	24	c ccagtaaatattggcaatccctaagaacggtttttagttgcacactcgggtgggaac	83						
Dbl	116028	CCCAGTAAATATTGGCAATGCCCGGACAFTTTTGGTTGTGCACACGGGAACAGAAC	116087						
Qy	84	atctactggcatctaaatgcataagaggcgcaagtatcgtctaacactctttcaaagcac	143						
Dbl	116088	CTGTC TACTGGCATCTAATG TAGTAGTGTAAGAATGCTGCCTGACA TCCTCACAGTGCACA	116147						
Qy	144	ggacagagcccacaaaagaanaattctagccccaaatgccataaacactcgtcttgag	203						
Dbl	116148	GGTCAGAACCCCATA CAAA GAATATCCAGCCCCAAATATCA CTAGTCGCGAGGCTGAG	116207						
Qy	204	aaacctaccgcaggatcttactgggttcataagtgtaagcttgccttgtttcgggttc	262						
Dbl	116208	AACACTGTTCTNAGGCCATTGCCAAATACCATGAGAGT GATGCTGTGAAGACTGGTTTC	116266						

RESULT	5	DNA	PRI	30-SEP-2000
LOCUS	AC007381/c	152081 bp	from 2,	complete sequence.
DEFINITION	Homo sapiens BAC clone RP11-158121			
ACCESSION	AC007381			
VERSION	AC007381.3			
KEYWORDS	HTG.			
SOURCE	human.			

SOURCE	ORGANISM	REFERENCE	TITLE	JOURNAL	REFERENCE	TITLE	JOURNAL	REFERENCE	TITLE	JOURNAL
human.	Homo sapiens	1 (bases 1 to 152081)	Sulston, J.E. and Waterston, R.	Toward a complete human genome sequence	Genome Res. 8 (11), 1097-1108 (1998)	99063792	2 (bases 1 to 152081)	Kozlowicz, A., Wohldmann, P. and Harper, M.	The sequence of Homo sapiens BAC clone RP11-158121	Unpublished
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Eutheria; Primates; Catarrhini; Hominidae; Homo.	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	1 (bases 1 to 152081)	Sulston, J.E. and Waterston, R.	Toward a complete human genome sequence	Genome Res. 8 (11), 1097-1108 (1998)	99063792	2 (bases 1 to 152081)	Kozlowicz, A., Wohldmann, P. and Harper, M.	The sequence of Homo sapiens BAC clone RP11-158121	Unpublished
Waterston, R.H.	Waterston, R.H.	Direct Submission	Submitted (25-APR-1999)	Genome Sequencing Center, Washington						

Wed Nov 7 09:21:19 2001

park parkway, St. Louis,

UNIVERSITY OF
MO 63108, USA
A (bases 1 to 152081)

REFERENCE	AUTHORS	TITLE	JOURNAL
-----------	---------	-------	---------

Waterson, R.H.
Direct Submission
Submitted (19-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

MO 63108, USA
5 (bases 1 to 152081)

REFERENCE	AUTHORS	TITLE	JOURNAL
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

Waterston, R.N.
Direct Submission
Submitted (21-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis
MO 63109 USA

MO 63108, USA
6 (bases 1 to 152081)

REFERENCE	AUTHORS	TITLE	JOURNAL
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

water, soil, and air.
Direct Submission
Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Dec 20, 1999 this sequence version replaced g1:5001522.

COMMENT

Center: Washington University genome sequencing
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@wustl.edu
 Summary Statistics
 Center project name: H_NH0158121

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping portions. It may be longer because we provide a small overlap between sections once, or longer because we provide a small overlap between sections once, or longer because we provide a small overlap between sections once, or longer because we provide a small overlap between sections once.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPEC1-11 human BAC library was made from the blood of one male donor, as described by Oseigawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateo, M., Catanese, J. D. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (pieter.dejong@roswellpark.edu)

VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP11-119H15.
The clone sequenced to the left is RP11-158I21.

The sequence Rp11-1581721 contains a dinucleotide (CT) repeat from base positions 6270 to 7430 in which the exact length is unknown. A PCR product from clone DNA confirms the approximate size of the region with band sizes of 2527 real, 2508 insilico for EcoRV; and 2527 real, 2508 insilico for HindIII.

FEATURES

```
1. 152081
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="2"
   /map="2"
```

[illegible]

us-09-656-668-198.rge

Wed Nov 7 09:21:19 2001

```

Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: DJ984123
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 93681 bases at least Q40
Consensus quality: 93706 bases at least Q30
Consensus quality: 93713 bases at least Q20
Insert size: 93714; sum-of-contigs
Insert size: 109355; 1.0% error; agarose-fp
Quality coverage: 8.25x in Q20 bases; sum-of-contigs Quality
coverage: 7.07x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
-----
FEATURES             Location/Qualifiers
     source            1..93714
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="20"
                        /clone_lib="RPCI-5"
                        /clone="RP5-984123"
     misc_feature       1..93714
                        /note="assembly_fragment:00029"
BASE COUNT            24148 a 22562 c 22969 g 24035 t
ORIGIN
Query Match          15.8%; Score 95.6; DB 79; Length 93714;
Best Local Similarity 75.8%; Pred. No. 2.9e-17;
Matches 144; Conservative 0; Mismatches 44; Indels 2; Gaps 2;
QY 20 ccatccaggaattgcaattctctaaagacgtgtttgtattgtcaccctgggtggg 79
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44680 CCACCCAGGGGACATTTACATGACTAGAGATGT-TTCTGGTTGTTCACAACTGGGT-GG 44623
QY 80 gaacatgctactgctatctaatgcataagggcagtaagtcgtgctaaacatcttcaacg 139
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44622 GAGGTGCTGCTGGCAGTCATGAGTGGTAGAGCCAGGAATGCTGCTAACATCTTACAATG 44563
QY 140 cacaggacagagcccccacaaagaataattctagcccccaaatgtccataacactgtgt 199
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44562 CCGGGACAGAGCCCTCCACGACAAAGAATTTATCCAGCCCCCAATGTCAATAGAGCTGAGGT 44503
QY 200 tgagaaaaacc 209
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44502 TGAGAAACCC 44493

RESULT 7
HS998H6 147815 bp DNA PRI 04-APR-2001
LOCUS Human DNA sequence from clone RP5-998H6 on chromosome 20q13.1.
DEFINITION Contains the gene for the ortholog of rat PB-Cadherin, ESI5, STSs,
GSS, two CpG islands and genomic marker D20S17, complete sequence.
ACCESSION AL031687
VERSION AL031687.1 GI:13366276
KEYWORDS HTG; CpG island; D20S17; PB-Cadherin.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 147815)
REFERENCE
AUTHORS Laird,G.
TITLE Direct Submission
JOURNAL Submitted (27-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
requests: clonerequest@sanger.ac.uk
----- Genome Center

/rpt_family="GA-rich"
13467..13930
/note="similar to Danio rerio EST AI958657 (NID:g5751370)"
fc94g02.y1
13470..13843
/note="similar to Mus musculus EST AI060775 (NID:g3336198)"
ub43e04.r1
13728..14181
/note="similar to Danio rerio EST AW116140 (NID:g6082478)"
fi09f10.x1
14169..14189
/rpt_family="(CAAA)n"
14238..14291
/rpt_family="(T)n"
14240..14639
/note="similar to Rattus norvegicus EST AA964818
(NID:g4279692)"
14249..14403
/note="similar to Rattus norvegicus EST AI704282
(NID:g4992182)"
14266..14451
/note="match to EST AA504508 (NID:g2240668) aa60g08.sl"
14287..14414
/note="similar to Danio rerio EST AW116140 (NID:g6082478)"
fi09f10.x1
14405..14886
/note="similar to Rattus sp. EST AW143415 (NID:g6163418)"
14497..14526
/rpt_family="(GA)n"
14541..14909
/note="similar to Rattus norvegicus EST AI029866
(NID:g4301527)"

Query Match          16.0%; Score 97.2; DB 86; Length 152081;
Best Local Similarity 72.2%; Pred. No. 1e-17;
Matches 140; Conservative 0; Mismatches 53; Indels 1; Gaps 1;
QY 17 ccccatcccgtaattgcaattctctaaagacgtgtttgtattgtcaccctgggt 76
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70146 CTCACCTCCCAAGGACATTTGGCAATGCTAGAGACAT-TTTTGGTTGTTCACAACTGGG 70088
QY 77 ggggaacatgctactggcatctaatgcataagggcagtaagtcgtgctaaacatctttca 136
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70087 GAGGGGGTGATCTAGTGGCATCTAGTAGAGAGCCCGGATGCTGCTAAATCTCTATA 70028
QY 137 acgcacagacagagcccccacaaagaataattctagcccccaaatgtccataacactgc 196
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70027 ATGCACAGACAGACACCCACAAAGAATTTCTCTCCCAAAATGTCAACAGTGCAGA 69968
QY 197 tggtagaaaaacct 210
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69967 GGTTAAGAACCTCT 69954

RESULT 6
AL161661/c 93714 bp DNA HTG 20-JAN-2001
LOCUS Homo sapiens chromosome 20 clone RP5-984123, *** SEQUENCING IN
DEFINITION PROGRESS ***, in unordered pieces.
ACCESSION AL161661
VERSION AL161661.1 GI:7263577
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 93714)
REFERENCE
AUTHORS Burton,J
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
----- Genome Center

```

us-09-656-668-198.rge

Wed Nov 7 09:21:19 2001

CB10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk

On Mar 16, 2001 this sequence version replaced gi:10198608. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/Celegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>

RP5-998H6 is from the library RPCI-5 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2

This sequence is the entire insert of clone RP5-998H6. The true right end of clone RP4-599F21 is at 29571 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Location/Qualifiers

1. .147815
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="q13.1"
/clone="RP5-998H6"
/clone_lib="RPCI-5"
85. .149
/note="MIR repeat: matches 60. .125 of consensus"

repeat_region
865. .905
/note="MER91 repeat: matches 153. .195 of consensus"

repeat_region
1141. .1358
/note="MIR repeat: matches 27. .250 of consensus"

repeat_region
1959. .2058
/note="L2 repeat: matches 2625. .2740 of consensus"

repeat_region
2081. .2176
/note="L2 repeat: matches 2612. .2710 of consensus"

repeat_region
2303. .2442
/note="7 copies 20 mer 81% conserved"

repeat_region
3585. .3790
/note="MIR repeat: matches 20. .239 of consensus"

repeat_region
4704. .4749
/note="23 copies 2 mer gt 84% conserved"

misc_feature
complement(4990. .5548)
/note="match: GSS: Em:AQ602943"

repeat_region
5125. .5354
/note="115 copies 2 mer tg 62% conserved"

repeat_region
5151. .5390
/note="12 copies 20 mer 60% conserved"

repeat_region
5397. .5442
/note="23 copies 2 mer gt 84% conserved"

repeat_region
5645. .5830
/note="MLT1B repeat: matches 1. .110 of consensus"

repeat_region
5831. .6141
/note="AluSg repeat: matches 1. .313 of consensus"

repeat_region
6142. .6394
/note="MLT1B repeat: matches 110. .390 of consensus"

repeat_region
6855. .7055
/note="MIR repeat: matches 59. .260 of consensus"

repeat_region
7129. .7544
/note="L2 repeat: matches 1529. .1989 of consensus"

7601. .8077
/note="MLT1J repeat: matches 34. .514 of consensus"

repeat_region
8089. .8174
/note="L2 repeat: matches 2066. .2155 of consensus"

repeat_region
8183. .8364
/note="MER53 repeat: matches 1. .182 of consensus"

repeat_region
10532. .10665
/note="MIR repeat: matches 80. .220 of consensus"

repeat_region
10708. .10764
/note="MIR repeat: matches 193. .251 of consensus"

repeat_region
11506. .11687
/note="MIR repeat: matches 29. .223 of consensus"

repeat_region
11702. .11810
/note="MIR repeat: matches 82. .190 of consensus"

repeat_region
11813. .12111
/note="AluIo repeat: matches 1. .312 of consensus"

repeat_region
12113. .12306
/note="97 copies 2 mer aa 66% conserved"

repeat_region
12123. .12302
/note="9 copies 20 mer 69% conserved"

repeat_region
12593. .12701
/note="MIR repeat: matches 64. .177 of consensus"

repeat_region
13003. .13232
/note="L2 repeat: matches 2266. .2499 of consensus"

repeat_region
13372. .13455
/note="MIR repeat: matches 72. .161 of consensus"

repeat_region
13574. .13780
/note="MIR repeat: matches 56. .262 of consensus"

repeat_region
14279. .14479
/note="MIR repeat: matches 65. .262 of consensus"

polyA_site
complement(14281)
/gene="dJ998H6.1"

mRNA
27368. .27489, 39965. .40101, 50849. .51102, 53537. .53730, 57368. .57535, 68050. .68169, 81505. .81799, 91582. .92237)
/gene="dJ998H6.1"
/note="match: cDNAs: Em:D83348 Em:D83349 Em:AB019618
Em:D82029 Em:D42150 Em:AF007116 Em:D25290 Em:D31784
Em:E07383 Em:D21254 Em:E07382 Em:D21255 Em:L34056
Em:L33477 Em:L34057 Em:X77557 Em:D31963 Em:D21253
Em:E07381 Em:D42149 Em:U59325 Em:X85330 Em:AB000512
Em:AF002983 Em:Y13919 Em:AF035300"
/evidence="not_experimental"
/product="dJ998H6.1 (ortholog of rat PB-Cadherin)"
complement(14281. .92237)
/gene="dJ998H6.1"
complement(14299. .14304)
/gene="dJ998H6.1"
14968. .15565
/note="CpG island"
/evidence="not_experimental"
complement(join(15048. .15619, 18488. .18739, 27130. .27247, 27368. .27489, 39965. .40101, 50849. .51102, 53537. .53730, 57368. .57535, 68050. .68169, 81505. .81799, 91582. .91836))
/gene="dJ998H6.1"
/note="match: proteins: Tr:Q63315 Tr:Q63561 Tr:Q90763
Sw:Q13634 Sw:P79995 Tr:Q15066 Tr:Q90762 Sw:P55285
Sw:P55280 Sw:P7326 Tr:Q9J1W2 Sw:P55287 Sw:P55288
Tr:Q93319 Tr:Q9Z0M3 Tr:Q9WTP5 Tr:O43205 Tr:Q9ULB2
Sw:P55286 Tr:O54800 Sw:P97291 Sw:P55289 Tr:Q9ULB5
Tr:Q9ULB4"
/codon_start=1
/evidence="not_experimental"
/product="dJ998H6.1 (ortholog of rat PB-Cadherin)"
/protein_id="CAB51587.2"
/db_xref="GI:13366277"
/db_xref="SPTREMBL:Q9UL99"
/translation="MRPRGGRAGRVAGVSPALLILLIPPTLLGLRWAGTPSP
SAGARQDGAIGRAGRWGNQFVVEYTGTEPLYVCKIHSDEGDEGAIKVTIS
GEAGATFLIDELTGDIHMERLDREQKFTYTLRAQDRATNRLLEPESEFTIKVQD
INDEPRFLHGFYIGSVAELSPGTSMQVMASDADPTYGSSARLYSVLGEHHFT
VDPKGTIVRTAVPDLDRSQREYVIVTQATDMAGQLGSGSTVTIVITVDVNDPPR
FPQKMYQFSIOESAPIGAVGKRVKAEVSDVGTDMVTYHLKDESSGGGVFKVYTDSD

* 170175 170274: gap of 100 bp
* 170175 170274: gap of 100 bp in length.

*	7402	8749:	contig of 1348 bp in length
*	8850	8849:	gap of 100 bp
*	8850	10630:	contig of 1781 bp in length
*	10631	10730:	gap of 100 bp
*	10731	12287:	contig of 1557 bp in length
*	12288	12387:	gap of 100 bp
*	12388	14291:	contig of 1904 bp in length
*	14292	14391:	gap of 100 bp
*	14392	15981:	contig of 1590 bp in length
*	15982	16081:	gap of 100 bp
*	16082	17733:	contig of 1292 bp in length
*	17374	17473:	gap of 100 bp
*	17474	19023:	contig of 1550 bp in length
*	19024	19123:	gap of 100 bp
*	19124	20568:	contig of 1445 bp in length
*	20569	20668:	gap of 100 bp
*	20669	22533:	contig of 1865 bp in length
*	22534	22633:	gap of 100 bp
*	22634	24930:	contig of 2297 bp in length
*	24931	25030:	gap of 100 bp
*	25031	27214:	contig of 2184 bp in length
*	27215	27314:	gap of 100 bp
*	27315	30058:	contig of 2744 bp in length
*	30059	30158:	gap of 100 bp
*	30159	31989:	contig of 1831 bp in length
*	31990	32089:	gap of 100 bp
*	32090	35144:	contig of 3055 bp in length
*	35145	35244:	gap of 100 bp
*	35245	37733:	contig of 2489 bp in length
*	37734	37833:	gap of 100 bp
*	37834	40275:	contig of 2442 bp in length
*	40276	40375:	gap of 100 bp
*	40376	42771:	contig of 2396 bp in length
*	42772	42871:	gap of 100 bp
*	42872	46769:	contig of 3898 bp in length
*	46770	46869:	gap of 100 bp
*	46870	49648:	contig of 2779 bp in length
*	49649	49748:	gap of 100 bp
*	49749	53828:	contig of 4080 bp in length
*	53829	53928:	gap of 100 bp
*	53929	56678:	contig of 4750 bp in length
*	56679	58778:	gap of 100 bp
*	58779	63068:	contig of 4290 bp in length
*	63069	63168:	gap of 100 bp
*	63169	67733:	contig of 4565 bp in length
*	67734	67833:	gap of 100 bp
*	67834	72886:	contig of 5053 bp in length
*	72887	72986:	gap of 100 bp
*	72987	76848:	contig of 3862 bp in length
*	76849	76948:	gap of 100 bp
*	76949	82756:	contig of 5808 bp in length
*	82757	82856:	gap of 100 bp
*	82857	88725:	contig of 5869 bp in length
*	88726	88825:	gap of 100 bp
*	88826	95518:	contig of 6693 bp in length
*	95519	95618:	gap of 100 bp
*	95619	101325:	contig of 5707 bp in length
*	101326	101425:	gap of 100 bp
*	101426	106203:	contig of 4778 bp in length
*	106204	106303:	gap of 100 bp
*	106304	111060:	contig of 4757 bp in length
*	111061	111160:	gap of 100 bp
*	111161	119653:	contig of 8493 bp in length
*	119654	119753:	gap of 100 bp
*	119754	127901:	contig of 8148 bp in length
*	127902	128001:	gap of 100 bp
*	128002	136682:	contig of 8681 bp in length
*	136683	136782:	gap of 100 bp
*	136783	146372:	contig of 9590 bp in length
*	146373	146472:	gap of 100 bp
*	146473	157184:	contig of 10712 bp in length
*	157185	157284:	gap of 100 bp
*	157285	170174:	contig of 12890 bp in length

FEATURES	Location/Qualifiers
source	1. .185143 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="RP11-766D10" /clone_lib="RPC1-11 Human Male BAC"
misc_feature	1. .200 /note="assembly_fragment clone_end:SP6 vector_side:left"
misc_feature	301. .1640 /note="assembly_fragment"
misc_feature	1741. .3026 /note="assembly_fragment"
misc_feature	3127. .4675 /note="assembly_fragment"
misc_feature	4776. .5775 /note="assembly_fragment"
misc_feature	5876. .7301 /note="assembly_fragment"
misc_feature	7402. .8749 /note="assembly_fragment"
misc_feature	8850. .10630 /note="assembly_fragment"
misc_feature	10731. .12287 /note="assembly_fragment"
misc_feature	12388. .14291 /note="assembly_fragment"
misc_feature	14392. .15981 /note="assembly_fragment"
misc_feature	16082. .17373 /note="assembly_fragment"
misc_feature	17474. .19023 /note="assembly_fragment"
misc_feature	19124. .20568 /note="assembly_fragment"
misc_feature	20669. .22533 /note="assembly_fragment"
misc_feature	22634. .24930 /note="assembly_fragment"
misc_feature	25031. .27214 /note="assembly_fragment"

Query Match	15.4%	Score 93.6;	DB 69;	Length 185143;
Best Local Similarity	66.2%;	Pred. No. 1.2e-16;		
Matches 135; Conservative	0;	Mismatches 69;	Indels	0;
Gaps	0;			

[illegible]

RESULT	9	HPG	12-MAR-2001
AL353791/c	208763 bp	DNA	
LOCUS	Human sapiens chromosome 9 clone RP11-475B17,		*** SEQUENCING IN
DEFINITION	PROGRESS ***, 4 unordered pieces.		
ACCESSION	AL353791		
VERSION	AL353791.5	GI:13373919	

Wed Nov 7 09:21:19 2001

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 208763)
Plumb, B.
Direct Submission
Submitted (11-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerquest@sanger.ac.uk
On Mar 16, 2001 this sequence version replaced gi:11340258.

COMMENT ----- Genome Center

Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bA472F14

----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 208161 bases at least Q40
Consensus quality: 208281 bases at least Q30
Consensus quality: 208368 bases at least Q20
Insert size: 208463; sum-of-contigs
Insert size: 209014; 4.8% error; agarose-fp
Quality coverage: 11.65x in Q20 bases; sum-of-contigs Quality
coverage: 11.81x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 119577: contig of 119577 bp in length
* 119578 119677: gap of 100 bp
* 119678 144219: contig of 24542 bp in length
* 144220 144319: gap of 100 bp
* 144320 204287: contig of 59968 bp in length
* 204288 204387: gap of 100 bp
* 204388 208763: contig of 4376 bp in length.

FEATURES Location/Qualifiers

Source
1..208763
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-475B17"
/clone_lib="RPC1-11.2"
1..119577
/note="assembly_fragment:00460
fragment_chain:1"
119678..144219
/note="assembly_fragment:00991
fragment_chain:1"
144320..204287
/note="assembly_fragment:04392
fragment_chain:1"
204388..208763
/note="assembly_fragment:05279.0"
60862 a 42636 c 42179 g 62780 t 306 others

BASE COUNT 60862 a 42636 c 42179 g 62780 t 306 others

ORIGIN

Query Match 15.4%; Score 93.6; DB 79; Length 208763;
Best Local Similarity 66.2%; Pred. No. 1.2e-16;
Matches 135; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 30 gaattattgcataagacgtgtttgtatgtcacacctgggtgggaacatgcta 89

Db 146724 GACATTTTGCATATCTGAGGCATTTTCAGTTATTATTAAGTGGGCTCGGTGATGCTG 146665
Qy 90 ctggcatctaataagagggcagtaataatgctgctaaacattttcaagcacaggacac 149
Db 146664 CTGGCATCTCATGAGAGAGACAGAGATGCTGTAAACATCTATATGACAGGACAG 146605
Qy 150 agcccccacaaagagaattatctagccccaaatgccataacactgctgttgagaaacc 209
Db 146604 GCCCCCAACACACAGATTATCCAGCCTAAATGTCACCTAGTCTGAGGTGAGAACTC 146545
Qy 210 taccgcagatcttactggccttc 233
Db 146544 TGACACATCGATTTCAGTGAATGTC 146521

RESULT 10

AC016951 147929 bp DNA PRI 05-MAY-2000
Homo sapiens 3 BAC RP11-452H12 (Roswell Park Cancer Institute Human
BAC Library) complete sequence.

AC016951
AC016951.9 GI:7658300

VERSION
KEYWORDS

HTG.

SOURCE

ORGANISM

Human.

Homo sapiens

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 147929)

Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K.,

Bodda, B., Bouck, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C.,

Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,

David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N.,

Duquand-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,

Forcum-Tansey, J., Frantz, P., Ganes, R., Garcia, D.K., Gorrell, J.H.,

Goirell, L.L., Guevara, W., Harris, K., Hosak, H., Jackson, L.E.,

Hodgson, A., Hoque, M., Holloway, C., Kondejewski, N., Kong, Y.,

Jackson, L., Jia, Y., Jones, M., Kelly, S., Kondejewski, N., Logan, O.,

Kovar, C., Leal, B., Li, Z., Lichtarge, O., Liu, J., Liu, W., McLeod, M.P.,

Lozano, R.J., Lu, J., Lucier, R., Martin, R., Martinez, C., Nash, S.,

Mei, G., Moore, S., Moorish, T., Morgan, M., Morris, S., Nash, S.,

Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S., Osval, G., Parish, B.,

Paxton, S., Payton, B., Perez, L., Pu, L.L., Quiles, M., Reiter, D.,

Rives, M., Samuel, S., Say, J., Scherer, S., Shah, E., Shen, H.,

Simon, M., Sparks, A., Stamps, A., Sucgang, R., Tabor, P., Taylor, T.,

Vasquez, L., Vinson, R., Vo, Q., Wabba, M., Watlington, S.,

Weinstock, G., Weinstock, I.R., Williamson, A., Worley, K., Wren, J.,

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.

entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics -----
Contig length: 147929
Phrap values in estimate: 146381
Average error rate (BCM-Phrap estimate): 0.000137522
Fraction of Phrap values less than 40 : 0.0255019
Number of consensus changing edits: 17
Number of N's in consensus : 0

Consensus changing edits		Edited+Context
Position	Original+Context	
26320	acctaagct(n)gatgacgagt	acctaagct(a)gatgacgagt
31259	ttcttttgg(n)aaagatgaac	ttcttttgg(g)aaagatgaac
37952	gaataatt(n)gaataaagat	gaataatt(g)gaataaagat
48020	agattcataa(n)gaagtcctg	agattcataa(a)gaagtcctg
61999	ttctaaaat(n)gtttgacctt	ttctaaaat(a)gtttgacctt
72566	tgattttgt(n)ccctggaagc	tgattttgt(t)ccctggaagc
81162	ctcataagc(n)taacatactg	ctcataagc(c)taacatactg
81177	atgaagtaag(n)caagccctg	atgaagtaag(a)caagccctg
82666	gcattggag(n)agatgggagg	gcattggag(g)agatgggagg
91636	aaattggaaa(n)aatccacact	aaattggaaa(c)aatccacact
92977	atgataggg(n)taacacaaatc	atgataggg(t)taacacaaatc
91637	cgctccatgc(n)caaaaaaaa	cgctccatgc(a)caaaaaaaa
96541	aaaaaaaatt(t)aaaaaaa	aaaaaaaatt(a)aaaaaaa
96574	aaaaaaaag(g)aaagatgtca	aaaaaaaag(a)aaagatgtca
98037	gctgcaagca(n)gccgggaagc	gctgcaagca(c)gccgggaagc
139118	gaggcagagg(n)caggaaagga	gaggcagagg(c)caggaaagga

----- Distribution of Quality < 40 Bases -----

1000	*	*
900	*	*
800	*	*
700	*	*
#		
bases	600	

Version: 1.01 qxf0.		Location/Qualifiers
FEATURES		source
500	*	1. .147929
400	*	/organism="Homo sapiens"
300	*	/db_xref="taxon:9606"
200	*	/chromosome="3"
100	*	/clone="RP11-452H12"
0	*	complement(532..617)
		/rpt_family="MSTB"
		625..696
		/rpt_family="LTR9"
		668..891
		/standard_name="Z53523"
		/db_xref="dbSTS:45636"
		697..731
		/rpt_family="(CA)n"
		complement(732..1090)
		/rpt_family="MSTB"
		1927..2384
		/rpt_family="MLT1G1"
		3079..3391
		/rpt_family="AluJb"
		complement(3685..3897)
		/rpt_family="MER20"
		complement(5392..5890)
		/rpt_family="Charliea"
		complement(5935..6618)
		/rpt_family="Charliea"
		8265..8469
		/standard_name="G09350"
		/db_xref="dbSTS:15731"
		8348..8390
		/rpt_family="(TTA)n"
		8503..8814
		/rpt_family="AluSx"
		9081..9226
		/rpt_family="MER53"
		9245..9413
		/rpt_family="MSTD"
		9414..9705
		/rpt_family="AluJb"
		9706..9951
		/rpt_family="MSTD"
		9952..10035
		/rpt_family="MER53"
		11657..13087
		/rpt_family="pTR5"
		complement(13237..13437)
		/rpt_family="MLT1H"
		19462..19885
		/rpt_family="HAL1"
		20035..20336
		/rpt_family="AluJb"
		20395..21174
		/rpt_family="L1PA2"
		21170..26420
		/rpt_family="L1PA2"
		25950..26082
		/standard_name="G19948"
		/db_xref="dbSTS:32826"

Phrap Value Range

5 10 15 20 25 30 35 40

Wed Nov 7 09:21:19 2001

vector_side:left*
BASE COUNT 59225 a 36880 c 38036 g 60172 t
ORIGIN
Query Match 15.1%; Score 91.8; DB 87; Length 147929;
Best Local Similarity 64.2%; Pred. No. 4e-16;
Matches 138; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
QY 2 gagttgccccctaccatccagtgaaatttgcaattctcctaagagcgtgtttga 61
Db 28127 GATTCTGGCACCACCCCAACCCAGGAAATTTTACATTTCTGGAGATATTTGGC 28186
QY 62 ttgtcaacacctgggtgggaacatgctacttgccatctaatcatagagggcgagtaact 121
Db 28187 TGTCCAACTTGTAGGAGAGGGGTGCTACTTGGCATCTTAATGAATAGAGGCCAAGGATCT 28246
QY 122 gctaaacatctttcaacgcagagagagcccccaaaaagaaattatctagcccaaa 181
Db 28247 GCTAAATCTTCTACATGCACAGGACAGCTCCGCCAACAAAGAAATTTTTCACACCAAAA 28306
QY 182 tgtccataacactgctgttgagaaacacctaccgca 216
Db 28307 TTTCACTAATCCAGAGGATGAGAAATACTGCTGTA 28341

RESULT 11
AL353729 194313 bp DNA HTG 06-MAR-2001
LOCUS Homo sapiens chromosome 9 clone RP11-290L7, *** SEQUENCING IN
DEFINITION PROGRESS ***, in unordered pieces.
ACCESSION AL353729 GI:113273646
VERSION AL353729.7
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 194313)
Plumb.B.
Direct Submission
Submitted (03-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Mar 11, 2001 this sequence version replaced gi:9800794.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA290L7
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: pGAP4; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 193402 bases at least Q40
Consensus quality: 193962 bases at least Q30
Consensus quality: 194209 bases at least Q20
Insert size: 194313; sum-of-contigs
Insert size: 196358; 3.1% error; agarose-fp
Quality coverage: 6.31x in Q20 bases; sum-of-contigs Quality
coverage: 6.45x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
Location/Qualifiers
1. .194313
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-290L7"
/clone_lib="RPCI-11.2"
1. .194313
/note="assembly_fragment:01008
clone_end:SP6

RESULT 12
AC017108 207746 bp DNA HTG 02-SEP-2000
LOCUS Homo sapiens chromosome UL clone RP11-566J10, WORKING DRAFT
DEFINITION SEQUENCE, 18 unordered pieces.
ACCESSION AC017108 GI:9965023
VERSION AC017108.4
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 207746)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 207746)
Waterston,R.H.
Direct Submission
Submitted (09-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Sep 2, 2000 this sequence version replaced gi:9838278.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information
Center project name: H.NH0566J10
----- Summary Statistics
Sequencing vector: pGAP4; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-terminator Big Dye; 17% of reads
Chemistry: Dye-terminator Big Dye; 17% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 191438 bases at least Q40
Consensus quality: 197227 bases at least Q30
Consensus quality: 200927 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 206046; sum-of-contigs
Quality coverage: 4.36 in Q20 bases; agarose-fp
Quality coverage: 3.63 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is

FEATURES
Source
1. .194313
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-290L7"
/clone_lib="RPCI-11.2"
1. .194313
/note="assembly_fragment:01008
clone_end:SP6

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 3206: contig of 3206 bp in length
 * 3207: gap of unknown length
 * 3307: contig of 8451 bp in length
 * 11757: gap of unknown length
 * 11858: contig of 13077 bp in length
 * 24934: gap of unknown length
 * 25034: contig of 11537 bp in length
 * 25035: gap of unknown length
 * 36571: gap of unknown length
 * 36572: contig of 14049 bp in length
 * 50720: gap of unknown length
 * 50721: contig of 10844 bp in length
 * 50821: gap of unknown length
 * 61664: gap of unknown length
 * 61665: contig of 15476 bp in length
 * 61765: gap of unknown length
 * 77241: gap of unknown length
 * 77341: contig of 27656 bp in length
 * 104997: gap of unknown length
 * 105097: contig of 27667 bp in length
 * 132763: gap of unknown length
 * 132863: contig of 27317 bp in length
 * 132764: gap of unknown length
 * 132864: contig of 4642 bp in length
 * 160181: gap of unknown length
 * 160280: contig of unknown length
 * 164922: gap of unknown length
 * 164923: contig of 1905 bp in length
 * 165023: gap of unknown length
 * 166927: contig of 3197 bp in length
 * 166928: gap of unknown length
 * 167028: contig of 3615 bp in length
 * 170224: gap of unknown length
 * 170225: contig of 3615 bp in length
 * 173939: gap of unknown length
 * 173940: contig of 7062 bp in length
 * 174040: gap of unknown length
 * 181101: contig of 8178 bp in length
 * 181201: gap of unknown length
 * 181202: contig of 9706 bp in length
 * 189380: gap of unknown length
 * 189479: contig of 9706 bp in length
 * 192885: gap of unknown length
 * 192886: contig of 8461 bp in length.
 * 199286: Location/Qualifiers

FEATURES

source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="UL"
 /clone="RP11-566J10"
 1. 3206
 /note="assembly_name:Contig1"
 3307. 11757
 /note="assembly_name:Contig10"
 11858. 24934
 /note="assembly_name:Contig11"
 25035. 36571
 /note="assembly_name:Contig12"
 clone_end:T7
 vector_side:right
 36672. 50720
 /note="assembly_name:Contig13"
 50821. 61664
 /note="assembly_name:Contig14"
 clone_end:SP6
 vector_side:left
 61765. 77240
 /note="assembly_name:Contig15"
 77341. 104996
 /note="assembly_name:Contig16"
 105097. 132763
 /note="assembly_name:Contig17"
 132864. 160180
 /note="assembly_name:Contig18"
 160281. 164922
 /note="assembly_name:Contig19"
 165023. 166927

/note="assembly_name:Contig3"
 167028. 170224
 /note="assembly_name:Contig4"
 170325. 173939
 /note="assembly_name:Contig5"
 174040. 181101
 /note="assembly_name:Contig6"
 181202. 189379
 /note="assembly_name:Contig7"
 189480. 199185
 /note="assembly_name:Contig8"
 199286. 207746
 /note="assembly_name:Contig9"
 60979 a 43162 c 40763 g 61116 t 1726 others
 BASE COUNT
 ORIGIN

Query Match 15.0%; Score 91; DB 64; Length 207746;
 Best Local Similarity 66.3%; Pred. No. 7.1e-16;
 Matches 130; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
 QY 24 cccagtgaaatttgcaattccaaagacgtgtttttgattgtcacaccctgggtgggaac 83
 Db 87414 CCCAGGACATGTGGCAATGTTGGAGACACTTTAGTGTGCACAACTTGGGGAGGTG 87355
 QY 84 atgtactggcatctaatgcatagagggcagtaagtctgctaaacatcttcaacgcaca 143
 Db 87354 GTGCTACTGGCAATTTAGTGTAGGTAGAGGTGAGAGACACTTCTAAACATCTCGTCAATGCACA 87295
 QY 144 ggcagagccccacaaagagaatattctagccccaaatgtccataacactgtgttgag 203
 Db 87294 GGACGACACCCACACAAAGAAATATCTCTCTAAATGACAAATGTGCTGAGACTGAG 87235
 QY 204 aaacactaccgcagga 219
 Db 87234 AGCCCTGCTTTAGGA 87219

RESULT 13
 AC021051 162082 bp DNA HTG 04-NOV-2000
 LOCUS Homo sapiens chromosome 3 clone RP11-214M2, WORKING DRAFT SEQUENCE,
 DEFINITION 6 unordered pieces.
 AC021051
 VERSION AC021051.9 GI:11079339
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 1 (bases 1 to 162082)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
 Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowler,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Denn,A.L., Ding,X., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Elhaj,C., Escotto,M., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Foster,P., Frantz,P., Gabel,J., Guevara,W., Gunaratne,P., Hale,S.,
 Garza,N., Gill,R., Gorrell,J.H., Harris,C., Hart,M., Havigak,P., Haves,A.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
 Hollins,B., Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,X., Johnson,R., Jolivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
 Loulseghe,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,

us-09-656-668-198.rge

wed Nov 7 09:21:19 2001

```
* 42884 42983: gap of 100 bp
* 42984 43690: contig of 707 bp in length
* 43691 43790: gap of 100 bp
* 43791 44484: contig of 694 bp in length
* 44485 44584: gap of 100 bp
* 44585 45265: contig of 681 bp in length
* 45266 45365: gap of 100 bp
* 45366 46078: contig of 713 bp in length
* 46079 46178: gap of 100 bp
* 46179 46886: contig of 708 bp in length
* 46887 46986: gap of 100 bp
* 46987 47700: contig of 714 bp in length
* 47701 47800: gap of 100 bp
* 47801 48505: contig of 705 bp in length
* 48506 48605: gap of 100 bp
* 48606 49293: contig of 688 bp in length
* 49294 49393: gap of 100 bp
* 49394 50141: contig of 748 bp in length
* 50142 50241: gap of 100 bp
* 50242 50943: contig of 702 bp in length
* 50944 51043: gap of 100 bp
* 51044 51745: contig of 702 bp in length
* 51746 51845: gap of 100 bp
* 51846 52554: contig of 709 bp in length
* 52555 52654: gap of 100 bp
* 52655 53348: contig of 694 bp in length
* 53349 53448: gap of 100 bp
* 53449 54172: contig of 724 bp in length
* 54173 54272: gap of 100 bp
* 54273 54957: contig of 685 bp in length
* 54958 55057: gap of 100 bp
* 55058 55758: contig of 701 bp in length
* 55759 55858: gap of 100 bp
* 55859 56570: contig of 712 bp in length
* 56571 56670: gap of 100 bp
* 56671 57378: contig of 708 bp in length
* 57379 57478: gap of 100 bp

Query Match      14.9%; Score 90.4; DB 67; Length 63020;
Best Local Similarity 66.3%; Pred. No. 9.5e-16;
Matches 130; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 24 cccagtgatatttgaattcctaagaacgtgttttgatgtcacacacctgggtggggaac 83
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 45912 CCCAGGGACATGTGGCAATGTTGGAGACACTTTAGGTGTGCACAACTTGGGGAGGTG 45853

Qy 84 atgctactggcatctaatgcatagagggcagtaatgctgtaaacatctttcaacgcaca 143
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 45852 GTGCTACTGGCATTATTAGTAGTAGGTCAGAGACACTGCTAAACATCTGCAATGCACA 45793

Qy 144 gacacagagcccaacaaagagaattatctagcccaaatgtccataacactgctgtgag 203
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 45792 GGCACAGACCCCAACACAAAGAATTATCTGTCTAAATGACAAATTGCTGCTGAGACTGAG 45733

Qy 204 aaacctaccgcaggga 219
      || | | | | | | |
Db 45732 AAGCCCTGCTTTAGGA 45717
```

Search completed: November 5, 2001, 23:02:50
Job time: 2800 sec

us-09-656-668-198.oli.rge

Wed Nov 7 09:21:18 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 5, 2001, 22:57:50 ; Search time 1188.78 Seconds
(without alignments)
7884.937 Million cell updates/sec

Title: US-09-656-668-198
Perfect score: 606
Sequence: 1 tgagtggcccttacccttcccc.....aagcctgtctgtctgcac 606

Scoring table: OLIGO-NUC
Gapop 60.0 , Gapext 60.0
Searched: 1344157 seqs, 7733874588 residues

Word size : 0
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_ba3.*
4: gb_in1.*
5: gb_in2.*
6: gb_in3.*
7: gb_om.*
8: gb_ov.*
9: gb_pat1.*
10: gb_pat2.*
11: gb_ph.*
12: gb_pl1.*
13: gb_pl2.*
14: gb_pl3.*
15: gb_pl4.*
16: em_ba1.*
17: em_ba2.*
18: em_fun.*
19: em_htgo_hum.*
20: em_htgo_inv.*
21: em_htgo_rod.*
22: em_htg_hum1.*
23: em_htg_hum2.*
24: em_htg_hum3.*
25: em_htg_hum4.*
26: em_htg_hum5.*
27: em_htg_hum6.*
28: em_htg_hum7.*
29: em_htg_hum8.*
30: em_htg_inv1.*
31: em_htg_inv2.*
32: em_htg_inv3.*
33: em_htg_rod.*
34: em_hum1.*
35: em_hum2.*
36: em_hum3.*
37: em_hum4.*
38: em_hum5.*
39: em_hum6.*
40: em_hum7.*
41: em_in.*
42: em_om.*
43: em_or.*

44: em_ov.*
45: em_pat.*
46: em_ph.*
47: em_pl.*
48: em_ro.*
49: em_sts.*
50: em_sy.*
51: em_un.*
52: em_v1.*
53: gb_sts1.*
54: gb_sts2.*
55: gb_sts3.*
56: gb_sy.*
57: gb_un.*
58: gb_v11.*
59: gb_v12.*
60: gb_htg1.*
61: gb_htg2.*
62: gb_htg3.*
63: gb_htg4.*
64: gb_htg5.*
65: gb_htg6.*
66: gb_htg7.*
67: gb_htg8.*
68: gb_htg9.*
69: gb_htg10.*
70: gb_htg11.*
71: gb_htg12.*
72: gb_htg13.*
73: gb_htg14.*
74: gb_htg15.*
75: gb_htg16.*
76: gb_htg17.*
77: gb_htg18.*
78: gb_htg19.*
79: gb_htg20.*
80: gb_htg21.*
81: gb_htg22.*
82: gb_htg23.*
83: gb_htg24.*
84: gb_htg25.*
85: gb_pr1.*
86: gb_pr2.*
87: gb_pr3.*
88: gb_pr4.*
89: gb_pr5.*
90: gb_pr6.*
91: gb_pr7.*
92: gb_pr8.*
93: gb_pr9.*
94: gb_rod.*
95: gb_rod2.*
96: gb_in4.*
97: gb_pr10.*
98: em_ba3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	606	100.0	606	10	AX093380	AX093380 Sequence
2	59	9.7	90497	92	HS732E4	AL008722 Human DNA
3	26	4.3	105137	93	HS1104017	AL080316 Human DNA
4	25	4.1	154969	62	AC012443	AC012443 Homo sapi
5	25	4.1	183893	77	AC090051	AC090051 Homo sapi
6	25	4.1	194294	76	AC079344	AC079344 Homo sapi
7	23	3.8	67084	78	AC090599	AC090599 Homo sapi
8	23	3.8	93959	92	HS106C24	283313 Human DNA S

C	9	23	3.8 103757	75	AC073940	AC073940 Homo sapi
C	10	23	3.8 127590	85	AC002554	AC002554 Human Chr
C	11	23	3.8 134394	83	AC067916	AC067916 Homo sapi
C	12	23	3.8 151357	62	AC011573	AC011573 Homo sapi
C	13	23	3.8 157402	71	AC034161	AC034161 Homo sapi
C	14	23	3.8 159723	66	AC021862	AC021862 Homo sapi
C	15	23	3.8 171589	75	AC078953	AC078953 Homo sapi
C	16	23	3.8 172214	64	AC016155	AC016155 Homo sapi
C	17	23	3.8 172588	60	AC008590	AC008590 Homo sapi
C	18	23	3.8 173808	89	AC036206	AC036206 Homo sapi
C	19	23	3.8 174128	89	AL137139	AL137139 Human DNA
C	20	23	3.8 176092	73	AC068222	AC068222 Homo sapi
C	21	23	3.8 178028	79	AL355338	AL355338 Homo sapi
C	22	23	3.8 178328	65	AC019265	AC019265 Homo sapi
C	23	23	3.8 180316	86	AC007773	AC007773 Homo sapi
C	24	23	3.8 186292	72	AC064876	AC064876 Homo sapi
C	25	23	3.8 186699	77	AC090015	AC090015 Homo sapi
C	26	23	3.8 189079	72	AC053545	AC053545 Homo sapi
C	27	23	3.8 190009	78	AC090519	AC090519 Homo sapi
C	28	23	3.8 191060	68	AC021468	AC021468 Homo sapi
C	29	23	3.8 194947	62	AC024534	AC024534 Homo sapi
C	30	23	3.8 202721	68	AC012564	AC012564 Homo sapi
C	31	22	3.6 161837	86	AC008479	AC008479 Homo sapi
C	32	22	3.6 174380	64	AC016075	AC016075 Homo sapi
C	33	22	3.6 178042	68	AC023986	AC023986 Homo sapi
C	34	22	3.6 194058	86	AC007564	AC007564 Homo sapi
C	35	22	3.6 225635	60	AC007553	AC007553 Homo sapi
C	36	21	3.5 39540	85	AC000079	AC000079 Homo sapi
C	37	21	3.5 43461	6	CER03D7	Z46828 Caenorhabdi
C	38	21	3.5 43934	85	AC000068	AC000068 Homo sapi
C	39	21	3.5 90590	81	AC010629	AC010629 Homo sapi
C	40	21	3.5 130757	76	AC079584	AC079584 Homo sapi
C	41	21	3.5 137658	92	HS1022311	AL049765 Human DNA
C	42	21	3.5 139457	80	AL357652	AL357652 Homo sapi
C	43	21	3.5 140002	90	AL354985	AL354985 Human DNA
C	44	21	3.5 150803	81	AL450398	AL450398 Homo sapi
C	45	21	3.5 155970	73	AC068085	AC068085 Homo sapi

ALIGNMENTS

```

1
RESULT 1
AX093380
LOCUS 506 bp DNA 30-MAR-2001
Sequence 198 from Patent WO0118046.
ACCESSION AX093380
VERSION AX093380.1 GI:13509828
KEYWORDS
SOURCE
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Xu, J. and Stolk, J.A.
TITLE Ovarian tumor sequences and methods of use therefor
JOURNAL Patent: WO 0118046-A 198 15-MAR-2001;
CORIXA CORPORATION (US)
FEATURES
source
1..606
/organism="Homo sapiens"
/db_xref="taxon:9606"
175 a 147 c 123 g 161 t
BASE COUNT
ORIGIN
Query Match 100.0%; Score 606; DB 10; Length 606;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 606; Conservative 0; Mismatches 0; Indels 0; Gaps
0
1 tgagttgccccctaccccccatcccgtagaatttgcatttccctaaagcgtgttg 60
db 1 TGAGTTGCCCCCTACCCCCATCCCGTAGAATTTGCCAATTCCTAAAGACGCTTTG 60

```

together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/C.elegans/wormpep> CTA-732E4 is from the human BAC library described in U-J. Kim et al. (1996) Genomics 34, 213-218.

VECTOR: pBelOAC11

This sequence is the entire insert of clone CTA-732E4 The true end of clone RP11-541J16 is at 5510 in this sequence. The true end of this sequence.

FEATURES
SOURCE

source	Location/Qualifiers
1. .90497	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="22"	
/map="q12.1"	
/clone="CTA-732E4"	
/clone_lib="CIT978SK-A2"	
2. .631	
repeat_region	/note="L1M1 repeat: matches 994. .1640 of consensus"
repeat_region	632. .757
repeat_region	/note="AluJb repeat: matches 1. .126 of consensus"
repeat_region	758. .1385
repeat_region	/note="L1P repeat: matches 3548. .4173 of consensus"
repeat_region	1374. .1869
repeat_region	/note="L1P repeat: matches 3221. .3716 of consensus"
repeat_region	1952. .2076
repeat_region	/note="L1MD2 repeat: matches 5142. .5259 of consensus"
repeat_region	2760. .3005
repeat_region	/note="L1MD2 repeat: matches 5259. .5173 of consensus"
repeat_region	3418. .3477
repeat_region	/note="30 copies 2 mer ac 90 conserved"
repeat_region	3423. .3478
repeat_region	/note="14 copies 4 mer caca 94 conserved"
repeat_region	3670. .3742
repeat_region	/note="MER82 repeat: matches 580. .653 of consensus"
repeat_region	3763. .4128
repeat_region	/note="MER82 repeat: matches 1. .386 of consensus"
misc_feature	3879. .4353
repeat_region	/note="match: GSS: Em:AQ559531"
repeat_region	4364. .4569
repeat_region	/note="L2 repeat: matches 2535. .2750 of consensus"
repeat_region	4612. .4930
repeat_region	/note="AluJo repeat: matches 1. .307 of consensus"
repeat_region	4934. .5220
repeat_region	/note="AluXo repeat: matches 1. .287 of consensus"
repeat_region	5221. .5268
repeat_region	/note="16 copies 3 mer taa 97 conserved"
misc_feature	complement(5266. .5727)
misc_feature	/note="match: GSS: Em:AQ020870"
misc_feature	complement(5282. .5740)
repeat_region	/note="match: GSS: Em:AQ557814"
repeat_region	5417. .5589
repeat_region	/note="MIR repeat: matches 2. .167 of consensus"
repeat_region	5645. .5708
repeat_region	/note="32 copies 2 mer at 71 conserved"
misc_feature	6469. .6726
repeat_region	/note="match: STS: Em:G03694; match: STS: Em:G03694"
repeat_region	7573. .7749
repeat_region	/note="MIR repeat: matches 22. .207 of consensus"
repeat_region	10322. .16495
repeat_region	/note="L1PA2 repeat: matches 7. .6146 of consensus"
misc_feature	16265. .16787
repeat_region	/note="match: GSS: Em:AQ186300"
repeat_region	16926. .16967
repeat_region	/note="21 copies 2 mer aa 76 conserved"
repeat_region	17850. .17971
repeat_region	/note="AluJo repeat: matches 19. .147 of consensus"

mailto:magnac.med.buffalo.edu/

This sequence is the entire insert of clone RP1-1040L7. This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>.

FEATURES

```

1. .105137
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q16.1-16.3"
/clone_lib="RPC1-1"
/clone="RP1-1040I7"
/complement(372..811)
/note="match: GSS: Em:AQ006718"
1037..1080
/note="22 copies 2 mer ta 88% conserved"
repeat_region
1153..1456
/note="AluX repeat: matches 1..301 of consensus"
repeat_region
2365..2702
/note="L2 repeat: matches 2363..2750 of consensus"
repeat_region
3579..3733
/note="L2 repeat: matches 2041..2201 of consensus"
repeat_region
4961..5272
/note="AluY repeat: matches 1..311 of consensus"
5545..5648
/note="52 copies 2 mer ta 71% conserved"
repeat_region
7388..7668
/note="AluJo repeat: matches 2..283 of consensus"
7871..8066
/note="MER58A repeat: matches 10..224 of consensus"
/note="complement(10286..10857)"
/misc_feature
/note="match: GSS: Em:B82911"
11075..11346
/note="LIME3 repeat: matches 5657..5940 of consensus"
/complement(13059..13460)
/misc_feature
/note="match: STS: Em:G21604"
14002..14199
/note="MER58C repeat: matches 3..89 of consensus"
17080..17249
/note="L1MC1 repeat: matches 6165..6323 of consensus"
17425..17538
/note="MIR repeat: matches 131..245 of consensus"
/misc_feature
/complement(17580..17784)
/note="match: STS: Em:G43022"
18332..18525
/note="AluJo repeat: matches 87..295 of consensus"
18919..19338
/note="L2 repeat: matches 1868..2326 of consensus"
19344..19733
/note="MLT1B repeat: matches 12..389 of consensus"
20175..20286
/note="56 copies 2 mer at 64% conserved"
repeat_region
20512..20976
/note="L2 repeat: matches 2264..2750 of consensus"
21113..21217
/note="L2 repeat: matches 1233..1341 of consensus"
21331..21517
/note="LIMEC repeat: matches 1768..1953 of consensus"
21882..22020
/note="LIME repeat: matches 419..560 of consensus"
22615..22710
/note="48 copies 2 mer tt 62% conserved"
repeat_region
23360..23417
/note="29 copies 2 mer tt 69% conserved"
24769..25327
/note="HERV16 repeat: matches 6..548 of consensus"
25328..26196
/note="MER6 repeat: matches 1..864 of consensus"
26197..26818
/note="HERV16 repeat: matches 548..1219 of consensus"

```

```

42857. .42971
/notes="L2 repeat: matches 2540. .2659 of consensus"
34009. .45826
/notes="L1M1 repeat: matches 2469. .4924 of consensus"
46433. .46883
/notes="LTR33 repeat: matches 24. .472 of consensus"
/notes="LTR33 repeat: matches 24. .472 of consensus"
47129. .47240
/notes="28 copies 4 mer gaga 66 conserved"
47132. .47241
/notes="55 copies 2 mer ag 67 conserved"
47276. .47434
/notes="MER5A repeat: matches 4. .189 of consensus"
47698. .47991
/notes="AluSc repeat: matches 1. .299 of consensus"
48005. .48060
/notes="FLAM_A repeat: matches 81. .133 of consensus"
48061. .48341
/notes="AluSg1 repeat: matches 1. .293 of consensus"
48399. .48549
/notes="FRAM repeat: matches -5. .146 of consensus"
49002. .49450
/notes="match: STS: Em:R42525; match: STS: Em:G24402"
49916. .50351
/notes="match: GSS: Em:AQ805065"
50076. .50449
/notes="match: GSS: Em:AQ805065"

```

misc_feature

	Score	Pred. No.	DB	Length
ery Match	9.7%	59	92	0
Local similarity	100.0%	3.5e-23		0

st Local Similarity	0; Mismatches	0; Indels
atches	59: Conservative	

Genes 55, 586

528 cactggagatgaataatactttaatttgtaacatglaaacaccctccca

|||||
- - - - -

TTTGTAACTTAATTTGTAACATGTGAAAACATCTATAAACATCTACTG 20172

20114 CACTGGAGATGAAATATACATTCTTTT

JUL 3 1966

FILE NO.	DATE	TIME	BY	REMARKS
104017	15-FEB-2000	105137	h0	DNA
104017	15-FEB-2000	105137	h0	DNA

JS
HSJ10401/ 103137 bp
Human DNA sequence from clone RP1-104017 on chromosome 6q10.1-10.3

INITIATION
Human DNA sequence
Contains STSS and GSSs, complete sequence.

SESSION
AL080316
CONFIDENTIAL

AL080316.8 GI:5650657

WORDS
HTG.
human

Human.
Homo sapiens
Vertebrata: Euteleostomi;

ORGANISM

Mammalia; Eutheria; Primates; Calathina; Mammalia; Eutheria; Primates; Calathina

REFERENCE

1 (bases 1 to 103137)

Tracey, A.
Direct Submission
Cambridgeshire,

Submitted (14-FEB-2000) Sanger Centre, Hinxton, Cambridge, UK
humquerv@sanger.ac.uk Clone

CB10 1SA, UK. E-mail enquiries: humquery@anger.ac.uk

```
requests: clonerquestesanger.ucsb.edu  
replaced gi:5566542.  
this sequence replaced gi:5566542.  
criteria
```

On Jul 29, 1999 this sequence has been finished according to sequence map criteria. This sequence has been finished according to sequencing problems; all sequencing problems;

This sequence has been followed as follows. An attempt is made to resolve all sequencing problems but not necessarily within known

such as compressions and repeats, but not necessarily where the sequence elements (e.g. Alu). Where the

annotated human repeat sequence elements (e.g. the 'unsure' annotation) there is an annotation using the

sequence is ambiguous, there is an unambiguous feature key

feature key. During sequence assembly data is compared from overlapping clones

Where differences are found these are annotated as variations. Note that the overlapping clone name. Note that the

together with a note of the overlapping clone names. Entries may not be found in the sequence submission

variation annotation may not be found in the overlapping clone, as we submit sequences with

corresponding to the overlapping, only a small overlap as described above.

The following abbreviations are used to associate primary access points with their source databases:

numbers given in the feature table with their source and the wormpep information.

Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Pf:, PIRATE database can be found at

on the WORMPEP database can be found at www.sanger.ac.uk/projects/C_elegans/wormpep. RPL-104017 is

from the library RPCI-1 constructed at the Roswell Park Cancer

FROM THE RESEARCH INSTITUTE BY THE GROUP OF PIETER DE JONG. FOR FURTHER DETAILS

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

us-09-656-668-198.oli.rge

Wed Nov 7 09:21:18 2001

```

repeat_region 26964..27296
/Note="LRR16A repeat: matches 95. .450 of consensus"
27297..27643
/Note="HERV1 repeat: matches 3587. .3946 of consensus"
28215..28338
/Note="MSTA repeat: matches 284. .422 of consensus"
28589..29009
/Note="match: GSS: Em:AQ308663"
28617..28812
/Note="MSTA repeat: matches 1. .178 of consensus"
29010..29383
/Note="MLT1B repeat: matches 1. .388 of consensus"
30078..30297
/Note="LRR16A repeat: matches 15. .229 of consensus"
32881..33162
/Note="LIMC5 repeat: matches 7649. .7926 of consensus"
33433..33616
/Note="MIR repeat: matches 40. .260 of consensus"
complement(33617..33834)
/Note="match: GSS: Em:AQ240281"
33899..34036
/Note="MLT11 repeat: matches 269. .404 of consensus"
37901..38084
/Note="AluSg/x repeat: matches 83. .264 of consensus"
38096..38119
/Note="12 copies 2 mer ac 95% conserved"
38855..38952
/Note="MIR repeat: matches 85. .184 of consensus"
39820..40101
/Note="AluY repeat: matches 1. .299 of consensus"
40284..41588
/Note="LTPA7 repeat: matches 2138. .3447 of consensus"
41588..44294
/Note="LTPA7 repeat: matches 3441. .6142 of consensus"
45641..46728
/Note="LIME1 repeat: matches 4994. .6109 of consensus"
46737..47808
/Note="TIGER1 repeat: matches 1350. .2418 of consensus"
47815..48092
/Note="Charlieb repeat: matches 70. .386 of consensus"
48179..48335
/Note="LIPB1 repeat: matches 5999. .6155 of consensus"
48338..48369
/Note="16 copies 2 mer aa 93% conserved"
48370..48882
/Note="LIM4 repeat: matches 4272. .4790 of consensus"
48927..49175
/Note="LIMEC repeat: matches 2146. .2412 of consensus"
49180..49590
/Note="LIMD3 repeat: matches 7310. .7739 of consensus"
49630..49717
/Note="LIMC4 repeat: matches 7219. .7313 of consensus"
49732..50676
/Note="LIMEC repeat: matches 1514. .2152 of consensus"
50716..50836
/Note="LIMEC repeat: matches 1287. .1413 of consensus"
50849..51871
/Note="LIMC1 repeat: matches 5287. .6316 of consensus"
51874..52239
/Note="LIMEC repeat: matches 929. .1302 of consensus"
52268..52519
/Note="LIMEC repeat: matches 523. .780 of consensus"
53098..53206
/Note="Alu repeat: matches 197. .305 of consensus"
53208..53388
/Note="MIR repeat: matches 41. .231 of consensus"
53414..53830
/Note="LTPA8 repeat: matches 5746. .6163 of consensus"
54168..54264
/Note="MIR repeat: matches 2. .111 of consensus"
misc_feature complement(54484..55028)
/Note="match: GSS: Em:AQ318586"
55956..56266
repeat_region

repeat_region 56471..56682
/Note="AluX repeat: matches 1. .312 of consensus"
56744..56947
/Note="L2 repeat: matches 2246. .2448 of consensus"
57428..57463
/Note="18 copies 2 mer tt 100% conserved"
complement(57599..58119)
/Note="match: GSS: Em:AQ346372"
58770..59081
/Note="AluY repeat: matches 1. .309 of consensus"
59703..60000
/Note="AluY repeat: matches 1. .294 of consensus"
60141..60194
/Note="L2 repeat: matches 2652. .2706 of consensus"
61142..61188
/Note="L2 repeat: matches 2652. .2698 of consensus"
61806..61910
repeat_region

Query Match 4.3%; Score 26; DB 93; Length 105137;
Best Local Similarity 100.0%; Pred.No. 0.0013;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 tgctactggtctaatgcatagag 110
|||||
DB 66659 TGCTACTGCCATCTAATGCATAGAG 66684

RESULT 4
AC012443 154969 bp DNA HTG 03-APR-2001
LOCUS Homo sapiens chromosome UNK clone RP11-17G11, WORKING DRAFT
DEFINITION SEQUENCE, 2 unchromed pieces.
AC012443
AC012443.7 GI:13518224
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 154969)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 154969)
Waterston,R.H.
Direct Submission
Submitted (27-OCT-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Apr 3, 2001 this sequence version replaced gi:13446343.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0017G11
----- Summary Statistics -----
Sequencing vector: M13; 34%
Chemistry: Dye-terminator Big Dye; 53% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 153554 bases at least Q40
Consensus quality: 153754 bases at least Q30
Consensus quality: 153862 bases at least Q20
Insert size: 157000; agarose-ff
Insert size: 153970; sum-of-contigs
Quality coverage: 7.72 in Q20 bases; sum-of-contigs
Quality coverage: 8.02 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently

```

Wed Nov 7 09:21:18 2001

* consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 153970: contig of 153970 bp in length
 153971 154070: gap of unknown length
 154071 154969: contig of 899 bp in length.

FEATURES

Source

1. 154969
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10p11"
 /clone="RP11-17G11"

misc_feature

1. 153970
 /note="assembly_name:Contig23"
 clone_end:SP6
 vector_side:right"

misc_feature

154071..154969
 /note="assembly_name:Contig5"
 48393 a 31230 c 30041 g 44839 t 466 others

BASE COUNT

Query Match 4.1%; Score 25; DB 62; Length 154969;
 Best Local Similarity 100.0%; Pred. No. 0.0051;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 gtgggaacatgctactgcatccta 99

Db 13126 GTGGGAACATGCTACTGTCATCTA 13150

RESULT 5

AC090051 183893 bp DNA HTG 04-MAR-2001
 LOCUS Homo sapiens chromosome 12q clone RP11-467E3, WORKING DRAFT
 DEFINITION SEQUENCE, 28 unordered pieces.

AC090051.2 GI:13194926

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 183893)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amarantunga,H.C., Are,J.R., Banks,T., Barbara,J.,
 Benton,J., Blimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Denny,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
 Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
 Loulseghe,H., Lozano,R.J., Lu,X., Lucier,A., Lucindale,A.,
 Ma,J., Maheshwari,M., Mapua,P., Martin,R., Meador,M.,
 Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
 Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,

Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
 Nguyen,A., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
 Oqub,M., Okwuonu,G., Oraqunye,N., Oviado,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,K., Primus,E., Pu,L.I.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
 Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,
 Sisson,T., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Thomas,N.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Vinson,R.,
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,
 Walli,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
 Watlington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S.,
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 and Gibbs,R.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: HCVV

Center clone name: RP11-467E3

----- Summary Statistics

Sequencing vector: M13; L08821

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 163729 bases at least Q40

Consensus quality: 171892 bases at least Q30

Consensus quality: 175386 bases at least Q20

Estimated insert size: 174626; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 3.8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 28 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 25049: contig of 25049 bp in length

* 25149: gap of unknown length

* 25150: contig of 20217 bp in length

* 43366: gap of unknown length

* 43367: contig of 19764 bp in length

* 45467: gap of unknown length

* 65231: contig of 9949 bp in length

* 75279: contig of 9940 bp in length

* 75280: contig of 10339 bp in length

* 85320: gap of unknown length

* 95759: contig of 9394 bp in length

* 105252: contig of 9779 bp in length

* 105253: contig of 7935 bp in length

* 115231: gap of unknown length

* 115232: contig of 7188 bp in length

* 123167: contig of 7188 bp in length

* 123267: gap of unknown length

* 130455: contig of 4057 bp in length

* 130555: contig of 4057 bp in length

FEATURES	source
134711: gap of unknown length	
134712: contig of 3497 bp in length	
138208: gap of unknown length	
138308: contig of 4201 bp in length	
142509: gap of unknown length	
142609: contig of 3207 bp in length	
142610: gap of unknown length	
145817: contig of 2657 bp in length	
145917: gap of unknown length	
148574: contig of 3911 bp in length	
148674: gap of unknown length	
152684: contig of 3621 bp in length	
156305: gap of unknown length	
156306: contig of 3307 bp in length	
156406: gap of unknown length	
159713: contig of 1964 bp in length	
159813: gap of unknown length	
161777: contig of 3184 bp in length	
161778: gap of unknown length	
165060: contig of 3100 bp in length	
165160: gap of unknown length	
165361: contig of 3171 bp in length	
168260: gap of unknown length	
168360: contig of 2394 bp in length	
168361: gap of unknown length	
171532: contig of 2372 bp in length	
171631: gap of unknown length	
171632: contig of 1081 bp in length	
174025: gap of unknown length	
174026: contig of 2773 bp in length	
174126: gap of unknown length	
176898: contig of 2131 bp in length	
176899: gap of unknown length	
179129: contig of 2372 bp in length	
179130: gap of unknown length	
179230: contig of 2372 bp in length	
181602: gap of unknown length	
181701: contig of 1081 bp in length	
182782: gap of unknown length	
182882: contig of 1011 bp in length	
182783: gap of unknown length	
182883: contig of 1011 bp in length	
Location/Qualifiers	
1. .183893	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="12q"	
/clone="RP11-467E3"	
54634 a 36554 c 36547 g 52928 t 3230 others	
BASE COUNT	
ORIGIN	
Query Match	4.1%; Score 25; DB 77; Length 183893;
Best Local Similarity	100.0%; Pred. No. 0.0051;
Matches	25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 86 gctactgcatctaatgcataagg 110	
Db 37859 GCTACTGTCATCTAATGCATAGAG 37883	
RESULT 6	
AC079344	194294 bp DNA HTG 31-AUG-2000
LOCUS	Homo sapiens chromosome UNK clone RP11-150L12, WORKING DRAFT
DEFINITION	SEQUENCE, 13 unordered pieces.
ACCESSION	AC079344
VERSION	AC079344.2 GI:9954828
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	Waterston,R.H.
JOURNAL	The sequence of Homo sapiens clone
AUTHORS	unpublished
TITLE	2 (bases 1 to 194294)
JOURNAL	Waterston,R.H.
AUTHORS	Direct Submission
TITLE	Submitted (28-AUG-2000) Genome Sequencing Center, Washington

Wed Nov 7 09:21:18 2001

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence.submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L11854
 Center clone name: 674_E_24

 * NOTE: This record contains 82 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 733 832: contig of 732 bp in length
 * 833 832: gap of 100 bp
 * 833 1536: contig of 704 bp in length
 * 1537 1636: gap of 100 bp
 * 1637 2355: contig of 719 bp in length
 * 2356 2455: gap of 100 bp
 * 2456 3173: contig of 718 bp in length
 * 3174 3273: gap of 100 bp
 * 3274 3969: contig of 696 bp in length
 * 3970 4069: gap of 100 bp
 * 4070 4793: contig of 724 bp in length
 * 4794 4893: gap of 100 bp
 * 4894 5626: contig of 733 bp in length
 * 5627 5736: gap of 100 bp
 * 5777 6445: contig of 719 bp in length
 * 6446 6545: gap of 100 bp
 * 6546 7262: contig of 717 bp in length
 * 7263 7362: gap of 100 bp
 * 7363 8088: contig of 726 bp in length
 * 8089 8188: gap of 100 bp
 * 8189 8905: contig of 717 bp in length
 * 8906 9005: gap of 100 bp
 * 9006 9725: contig of 720 bp in length
 * 9726 9825: gap of 100 bp
 * 9826 10539: contig of 714 bp in length
 * 10540 10639: gap of 100 bp
 * 10640 11354: contig of 715 bp in length
 * 11355 11454: gap of 100 bp
 * 11455 12164: contig of 710 bp in length
 * 12165 12264: gap of 100 bp
 * 12265 12974: contig of 710 bp in length
 * 12975 13074: gap of 100 bp
 * 13075 13806: contig of 732 bp in length
 * 13807 13906: gap of 100 bp
 * 13907 14628: contig of 722 bp in length
 * 14629 14728: gap of 100 bp
 * 14729 15460: contig of 732 bp in length
 * 15461 15560: gap of 100 bp
 * 15561 16277: contig of 717 bp in length
 * 16278 16377: gap of 100 bp
 * 16378 17115: contig of 738 bp in length
 * 17116 17215: gap of 100 bp
 * 17216 17922: contig of 707 bp in length
 * 17923 18022: gap of 100 bp
 * 18023 18732: contig of 710 bp in length
 * 18733 18832: gap of 100 bp
 * 18833 19546: contig of 714 bp in length
 * 19547 19646: gap of 100 bp
 * 20358: contig of 712 bp in length

misc_feature 13454..19171
 /note="assembly_name:Contig13"
 misc_feature 19272..32004
 /note="assembly_name:Contig14"
 misc_feature 32105..43844
 /note="assembly_name:Contig16"
 misc_feature 43945..53017
 /note="assembly_name:Contig17"
 misc_feature 53118..64048
 /note="assembly_name:Contig18"
 misc_feature 64149..77841
 /note="assembly_name:Contig19"
 misc_feature 77942..95865
 /note="assembly_name:Contig20"
 misc_feature 95966..122392
 /note="assembly_name:Contig21"
 misc_feature 122493..153546
 /note="assembly_name:Contig22"
 misc_feature 153647..194294
 /note="assembly_name:Contig23"
 BASE COUNT 60393 a 37283 c 37173 g 58245 t 1200 others
 ORIGIN
 Query Match 4.1%; Score 25; DB 76; Length 194294;
 Best Local Similarity 100.0%; Pred. No. 0.0052;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 75 gtgggaacatgctactggcatcta 99
 |||||||||||||||||||||||||
 DB 73447 GTGGGGAACATGCTACTGGCATCTA 73471
 RESULT 7
 AC090599/c 67084 bp DNA HTG 03-MAR-2001
 LOCUS Homo sapiens chromosome 11 clone RP11-674E24 map 11, LOW-PASS
 DEFINITION SEQUENCE SAMPLING.
 AC090599
 AC090599.1 GI:13194332
 VERSION
 KEYWORDS HTG; HTGS_PHASE0.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
 Birren,B., Bastien,V., Boguslavsky,L., Bouckgalter,B., Brown,A.,
 Barna,N., Bastien,V., Choepel,Y., Colangelo,M., Collins,S.,
 Canarata,J., Campopiano,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
 Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Fero,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
 Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-pierre,N.,
 Hagos,B., Headford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
 Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
 Lehotzky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
 Marquis,N., Mathews,C., McCarthy,M., McEwan,P., McKernan,K.,
 McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
 Phunhng,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
 Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
 Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S., Severy,P.,
 Souknez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Travers,M., Travis,N., Trigglio,J., Vassiliev,H., Viel,R., Vo,A.,
 Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
 Zembek,U., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (03-MAR-2001) Whitehead Institute/MIT Center for Genome

```

* 49945 50658: contig of 713 bp in length
* 50659 50758: gap of 100 bp
* 50759 51472: contig of 714 bp in length
* 51473 51572: gap of 100 bp
* 51573 52284: contig of 712 bp in length
* 52285 52384: gap of 100 bp
* 52385 53101: contig of 717 bp in length
* 53102 53201: gap of 100 bp
* 53202 53935: contig of 734 bp in length
* 53936 54035: gap of 100 bp
* 54036 54767: contig of 732 bp in length
* 54768 54867: gap of 100 bp
* 54868 55587: contig of 720 bp in length
* 55588 55687: gap of 100 bp
* 55688 56419: contig of 732 bp in length
* 56420 56519: gap of 100 bp

Query Match 3.84; Score 23; DB 78; Length 67084;
Best Local Similarity 100.0%; Pred. No. 0.075;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 agaattatctagcccaaatgctc 185
|||||
Db 33966 AGAATTATCTAGCCCCAAATGTC 33944

RESULT 8
LOCUS HS106C24 93959 bp DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from PAC 106C24, between markers DXS294 and
DSX730 on chromosome X.
ACCESSION 283313
VERSION 283313.1 GI:1730462
KEYWORDS X.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 93959)
Isheredwood,J.
Direct Submission
Submitted (02-DEC-1996) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
IMPORTANT: This sequence is not the entire insert of clone 106C24.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The true left end of clone 106C24 is at 1 in this sequence. The
true right end of clone 206J9 is at 47751.
106C24 is from a whole genome PAC library.
location/Qualifiers
source
1. .93959
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="X"
/clone_lib="RPC1-1"
/clone="rp1-106C24"
552. .799
/note="MIR repeat: matches 259. .3 of consensus"
repeat_region
2498. .2582
/note="MIR repeat: matches 57. .145 of consensus"
repeat_region
3438. .3504
/note="MIR repeat: matches 122. .56 of consensus"
repeat_region
4744. .4938
/note="MER3 repeat: matches 2. .200 of consensus"

```

```

* 20359 20458: gap of 100 bp
* 20459 21188: contig of 730 bp in length
* 21189 21288: gap of 100 bp
* 21289 22004: contig of 716 bp in length
* 22005 22104: gap of 100 bp
* 22105 22805: contig of 701 bp in length
* 22806 22905: gap of 100 bp
* 22906 23633: contig of 728 bp in length
* 23634 23733: gap of 100 bp
* 23734 24453: contig of 720 bp in length
* 24454 24553: gap of 100 bp
* 24554 25281: contig of 728 bp in length
* 25282 25381: gap of 100 bp
* 25382 26107: contig of 726 bp in length
* 26108 26207: gap of 100 bp
* 26208 26925: contig of 718 bp in length
* 26926 27025: gap of 100 bp
* 27026 27748: contig of 723 bp in length
* 27749 27848: gap of 100 bp
* 27849 28576: contig of 728 bp in length
* 28577 28676: gap of 100 bp
* 28677 29417: contig of 741 bp in length
* 29418 29517: gap of 100 bp
* 29518 30229: contig of 712 bp in length
* 30230 30329: gap of 100 bp
* 30330 31033: contig of 704 bp in length
* 31034 31133: gap of 100 bp
* 31134 31844: contig of 711 bp in length
* 31845 31944: gap of 100 bp
* 31945 32640: contig of 696 bp in length
* 32641 32740: gap of 100 bp
* 32741 33442: contig of 702 bp in length
* 33443 33542: gap of 100 bp
* 33543 34261: contig of 719 bp in length
* 34262 34361: gap of 100 bp
* 34362 35079: contig of 718 bp in length
* 35080 35179: gap of 100 bp
* 35180 35913: contig of 734 bp in length
* 35914 36013: gap of 100 bp
* 36014 36743: contig of 730 bp in length
* 36744 36843: gap of 100 bp
* 36844 37575: contig of 732 bp in length
* 37576 37675: gap of 100 bp
* 37676 38393: contig of 718 bp in length
* 38394 38493: gap of 100 bp
* 38494 39213: contig of 720 bp in length
* 39214 39313: gap of 100 bp
* 39314 40023: contig of 710 bp in length
* 40024 40123: gap of 100 bp
* 40124 40846: contig of 723 bp in length
* 40847 40946: gap of 100 bp
* 40947 41641: contig of 695 bp in length
* 41642 41741: gap of 100 bp
* 41742 42459: contig of 718 bp in length
* 42460 42559: gap of 100 bp
* 42560 43258: contig of 699 bp in length
* 43259 43358: gap of 100 bp
* 43359 44090: contig of 732 bp in length
* 44091 44190: gap of 100 bp
* 44191 44920: contig of 730 bp in length
* 44921 45020: gap of 100 bp
* 45021 45739: contig of 719 bp in length
* 45740 45839: gap of 100 bp
* 45840 46561: contig of 722 bp in length
* 46562 46661: gap of 100 bp
* 46662 47378: contig of 717 bp in length
* 47379 47478: gap of 100 bp
* 47479 48194: contig of 716 bp in length
* 48195 48294: gap of 100 bp
* 48295 49024: contig of 730 bp in length
* 49025 49124: gap of 100 bp
* 49125 49845: contig of 721 bp in length
* 49846 49945: gap of 100 bp

```

Wed Nov 7 09:21:18 2001

repeat_region	5365. .5599	/note="MIR repeat: matches 8. .236 of consensus"	repeat_region	/note="MIR repeat: matches 152. .29 of consensus"
repeat_region	7810. .7942	/note="MIR repeat: matches 13. .154 of consensus"	repeat_region	35255. .35321
repeat_region	8410. .8520	/note="MIR repeat: matches 11. .119 of consensus"	repeat_region	35621. .36097
repeat_region	9006. .9089	/note="MIR repeat: matches 63. .146 of consensus"	repeat_region	35688. .36721
repeat_region	9575. .9767	/note="MIR repeat: matches 5020. .4829 of consensus"	repeat_region	38850. .39742
repeat_region	10196. .10275	/note="MIR repeat: matches 106. .4 of consensus"	repeat_region	39598. .44765
repeat_region	10515. .10619	/note="MIR repeat: matches 49. .153 of consensus"	repeat_region	45104. .45334
repeat_region	11164. .11350	/note="MIR repeat: matches 702. .907 of consensus"	repeat_region	48462. .48649
repeat_region	12097. .12302	/note="MIR repeat: matches 30. .262 of consensus"	repeat_region	50125. .53643
repeat_region	12338. .13023	/note="MIR repeat: matches 146. .64 of consensus"	repeat_region	53494. .54383
repeat_region	14787. .14844	/note="MIR repeat: matches 29. .90 of consensus"	repeat_region	54743. .55195
repeat_region	17057. .17086	/note="MIR repeat: matches 10. .93 of consensus"	repeat_region	55530. .56049
repeat_region	17105. .17343	/note="MIR repeat: matches 258. .11 of consensus"	repeat_region	57142. .57399
repeat_region	17750. .18041	/note="MIR repeat: matches 297. .5 of consensus"	repeat_region	57251. .58278
repeat_region	19193. .19360	/note="MIR repeat: matches 213. .49 of consensus"	repeat_region	58905. .59344
repeat_region	20826. .20962	/note="MIR repeat: matches 1. .136 of consensus"	repeat_region	59359. .59409
repeat_region	20964. .21104	/note="MIR repeat: matches 92. .234 of consensus"	repeat_region	59440. .59664
repeat_region	21093. .21869	/note="MIR repeat: matches 5. .772 of consensus"	repeat_region	60853. .60918
repeat_region	21856. .26363	/note="MIR repeat: matches 887. .5390 of consensus"	repeat_region	61282. .61335
repeat_region	26214. .27104	/note="MIR repeat: matches 1. .891 of consensus"	repeat_region	61374. .61454
repeat_region	29529. .29582	/note="MIR repeat: matches 66. .120 of consensus"	repeat_region	61902. .62103
repeat_region	30103. .30321	/note="MIR repeat: matches 374. .156 of consensus"	repeat_region	64007. .64146
repeat_region	30339. .30589	/note="MIR repeat: matches 909. .659 of consensus"	repeat_region	64287. .64532
repeat_region	30596. .30743	/note="MIR repeat: matches 149. .1 of consensus"	repeat_region	67972. .68217
repeat_region	30834. .31103	/note="MIR repeat: matches 1231. .961 of consensus"	repeat_region	68347. .68471
repeat_region	31580. .31951	/note="MIR repeat: matches 599. .231 of consensus"	repeat_region	70707. .70748
repeat_region	31957. .32266	/note="MIR repeat: matches 297. .1 of consensus"	repeat_region	71075. .71375
repeat_region	3267. .32482	/note="MIR repeat: matches 249. .5 of consensus"	repeat_region	71608. .71649
repeat_region	32487. .32545	/note="MIR repeat: matches 374. .313 of consensus"	repeat_region	71736. .71944
repeat_region	32551. .32850	/note="MIR repeat: matches 1. .301 of consensus"	repeat_region	71985. .72402
repeat_region	32870. .32911	/note="MIR repeat: matches 270. .312 of consensus"	repeat_region	72407. .72502
repeat_region	32912. .33202	/note="MIR repeat: matches 297. .1 of consensus"	repeat_region	72503. .72804
repeat_region	33703. .33777	/note="MIR repeat: matches 345. .271 of consensus"	repeat_region	72806. .72876
repeat_region	33811. .33871	/note="MIR repeat: matches 990. .1050 of consensus"	repeat_region	72877. .72979
repeat_region	33898. .34115	/note="MIR repeat: matches 210. .1 of consensus"	repeat_region	72909. .73456
repeat_region	35085. .35213	/note="MIR repeat: matches 1231. .961 of consensus"	repeat_region	/note="MIR repeat: matches 32. .593 of consensus"

Wed Nov 7 09:21:18 2001

* of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved
 * 1 103757: contig of 103757 bp in length.

FEATURES

source
 1. 103757
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="15"
 /map="15q15"
 /clone="CTD-2190J1"
 /clone_lib="Cal Tech human BAC library D"
 /note="This clone overlaps RP11-213B23 and RP11-616K22"
 /note="21521 c 21252 g 28399 t"

BASE COUNT 32585 a 21521 c 21252 g 28399 t
 ORIGIN
 3.8%; Score 23; DB 75; Length 103757;
 Query Match 100.0%; Pred. No. 0.077;
 Best Local Similarity 0; Mismatches 0; Indels 0; Gaps 0;
 Matches 23; Conservative 0;

QY 163 agaattatctagcccaaatgctc 185
 |||||
 Db 94438 AGAATTATCTAGCCCAATGTC 94416

RESULT 10

AC002554 127590 bp DNA PRI 27-AUG-1998
 LOCUS Human Chromosome 1p11.2 PAC clone pDJ404m15, complete sequence.
 DEFINITION AC002554
 ACCESSION AC002554
 VERSION 1 GI:3478647
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 127590)
 Evans, G.A., Athanasiou, M., Aguayo, P., Armstrong, D., Basit, M.,
 Buettner, J., Bumeister, R., Card, P., deSailboat, F., Dunn, J.,
 English, C., Ethridge, S., Garner, H.R., Gee, V., Gordon, M., Gotway, G.,
 Grant, O., Hahner, L., Joslin, J., Lewis, E., Loo, H., Loo, K.N.,
 Major, T., McFarland, J., Newton, J., Osborne-Lawrence, S.,
 Schageman, J., Schultz, R.A., Stimson, S., Syed, M. and Ward, T.
 HTGS Submission

REFERENCE

2 (bases 1 to 127590)
 Evans, G.A., Athanasiou, M., Bradbury, P., Brignac, S., Bumeister, R.,
 Davis, C., English, C., Franklin, T.L., Garner, H.R., Gordon, M.,
 Gotway, G., Grant, O., Hahner, L., Harris, J., Hinson, S.,
 Narayanaswamy, U., Newton, J., O'Brien, K., Oliver, T., Patel, P.,
 Probst, S., Rayner, S., Schageman, J., Schilling, P., Schultz, R.,
 Syed, M., Valenzuela, D., Ward, T. and Wilson, R.
 Direct Submission
 Submitted (23-SEP-1997) Genome Science and Technology Center,
 University of Texas Southwestern Medical Center at Dallas, 5323
 Harry Hines Blvd, Dallas, TX 75235-8591, USA
 3 (bases 1 to 127590)
 Evans, G.A., Athanasiou, M., Basit, M., Bradbury, P., Brignac, S.,
 Bumeister, R., Davis, C., English, C., Franklin, T.L., Garner, H.R.,
 Gee, V., Gordon, M., Gotway, G., Grant, O., Hahner, L., Harris, J.,
 Hinson, S., Narayanaswamy, U., Newton, J., O'Brien, K., Patel, P.,
 Schageman, J., Schilling, P., Schultz, R., Syed, M., Valenzuela, D.,
 Ward, T. and Wilson, R.
 Direct Submission
 Submitted (17-DEC-1997) Genome Science & Technology Center,
 University of Texas Southwestern Medical Center, 5323 Harry Hines
 Blvd, Dallas, TX 75235-8591, USA
 4 (bases 1 to 127590)
 Evans, G.A., Athanasiou, M., Aguayo, P., Armstrong, D., Basit, M.,
 Buettner, J., Bumeister, R., Card, P., deSailboat, F., Dunn, J.,

TITLE

JOURNAL
 REFERENCE
 AUTHORS

REFERENCE

AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS

1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 Genbank flat file format but are available as part
 of this entry's ASN.1 file.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes

repeat_region 73468..73517
 /note="25 copies of 2 mer 96 & conserved"
 repeat_region 73543..74432
 /note="MSI-INTERNAL repeat: matches 532..1458 of
 consensus"
 repeat_region 74466..75346
 /note="11PAL2 repeat: matches 912..4 of consensus"
 repeat_region 75199..75954
 /note="L1 repeat: matches 5390..767 of consensus"

Query Match 3.8%; Score 23; DB 92; Length 93959;
 Best Local Similarity 100.0%; Pred. No. 0.076;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 agaattatctagcccaaatgctc 185
 |||||
 Db 85580 AGAATTATCTAGCCCAATGTC 85602

RESULT 9
 AC073940/c DNA HTG 14-APR-2001
 LOCUS Homo sapiens chromosome 15 clone CTD-2190J1 map 15q15, ***
 DEFINITION SEQUENCING IN PROGRESS ***, 1 ordered pieces.

AC073940
 VERSION AC073940.3 GI:13624386
 KEYWORDS HTG; HTGS_PHASE2; HTGS_FULLTOP.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 103757)
 Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
 Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
 Pate, D. and Hood, L.
 Sequencing of human chromosome 15 D15S146-D15S117 region
 Unpublished

REFERENCE

AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 On Apr 14, 2001 this sequence version replaced gi:12658005.

 Center: Genome Center
 Center code: UWMSC
 Web site: http://chroma.mbt.washington.edu/msg_www
 Contact: leetowen@systemsbiology.org

 Summary Statistics
 Sequencing vector: pUC18; L08752
 Chemistry: Dye-terminator Big Dye; 90% of reads
 Chemistry: Dye-terminator Big Dye; 10% of reads
 Assembly program: Phrap; version 0.990399
 Insert size: 100000; agarose-ff
 Quality coverage: 10.2x in Q20 bases; sum-of-contigs

COMMENT

Sequence Quality Assessment:
 This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 Genbank flat file format but are available as part
 of this entry's ASN.1 file.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes

us-09-656-668-198.oli.rge

Wed Nov 7 09:21:18 2001

English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Joslin,J., Lewis,E., Loo,H., Loo,K.N., Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.

Direct Submission
Submitted (27-AUG-1998) Genome Science & Technology Center,
University of Texas Southwestern Medical Center, 5323 Harry Hines
Bldg. Dallas, TX 75235-8591, USA

On Aug 27, 1998 this sequence version replaced gi:2695561.
region mapped between STS markers D11S578 and D11S3805h.
CHROMOSOMAL LOCUS: This PAC clone comes from the lip11.2
MARKER CONFIRMATION: STS sequence confirmed; D11S578

MAPPED CLONE OVERLAP: PDJ404c10.
IMPORTANT: This submission contains the entire insert of clone
PDJ404m15. PDJ404m15 comes from the RPI-3 PAC library constructed
at the Roswell Park Cancer Institute by the Pletier de Jong group.
This clone has been finished according to strict quality criteria
and attempts have been made to resolve all base calling problems
such as compressions and repetitive elements. The expected
Phred/Phrap calculated errors/10kb is 0.23. In addition, this
sequence has been finished such that 99.9% of consensus base calls
consist of either double-stranded coverage or 2 types of labeling
chemistry on one strand. Additional information regarding this
clone may be found on the GENE web page at
<http://genec.swned.edu/lip11.htm>.

FEATURES

Source 1. 127590 Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
complement(1. 69)
/rpt_family="MER41"
90. 385
/rpt_family="Alu"
complement(400. 821)
/rpt_family="MER41"
complement(1101. 1382)
/rpt_family="LFR9"
complement(1436. 1731)
/rpt_family="Alu"
complement(1797. 1937)
/rpt_family="LFR9"
complement(3274. 3558)
/rpt_family="Alu"
complement(3923. 4100)
/rpt_family="MER4"
4779. 5469
/rpt_family="LFR8"
complement(7262. 7513)
/rpt_family="Alu"
complement(7651. 7964)
/rpt_family="Alu"
11021. 11306
/rpt_family="Alu"
11424. 12625
/rpt_family="L1"
13008. 13297
/rpt_family="Alu"
13356. 13506
/rpt_family="L1"
complement(14032. 14334)
/rpt_family="Alu"
14861. 15068
/rpt_family="MER20"
15934. 16238
/rpt_family="Alu"
16914. 17018
/rpt_family="Alu"
complement(17093. 17337)
/rpt_family="Alu"
complement(17362. 17691)
/rpt_family="MER42"
17918. 18156
/rpt_family="Alu"

repeat_region complement(18206. 18402)
/rpt_family="MIR"
19740. 19807
/rpt_family="Alu"
complement(21285. 21395)
/rpt_family="MLT1"
21587. 21762
/rpt_family="MLT1"
21779. 21846
/rpt_family="MLT1"
21857. 21924
/rpt_family="MSTAR"
23972. 24546
/rpt_family="MLT1"
24638. 24733
/rpt_family="MIR"
complement(24860. 25148)
/rpt_family="Alu"
26031. 26203
/rpt_family="Alu"
26528. 26717
/rpt_family="Alu"
complement(27650. 27708)
/rpt_family="MIR"
complement(29082. 29343)
/rpt_family="Alu"
30684. 30980
/rpt_family="Alu"
32634. 32818
/rpt_family="Alu"
complement(35666. 35779)
/rpt_family="Alu"
36025. 36096
/rpt_family="MIR"
36923. 37030
/rpt_family="MIR"
complement(37101. 37189)
/rpt_family="THE1"
37653. 37940
/rpt_family="Alu"
38374. 38647
/rpt_family="Alu"
complement(38907. 39198)
/rpt_family="Alu"
complement(40840. 40945)
/rpt_family="MER5"
complement(41499. 41785)
/rpt_family="Alu"
42735. 43073
/rpt_family="MLT1"
complement(43666. 43951)
/rpt_family="Alu"
44899. 45060
/rpt_family="MIR"
complement(45617. 45784)
/rpt_family="Alu"
45878. 46030
/rpt_family="MIR"
complement(46105. 46157)
/rpt_family="MLT1"
complement(46177. 46305)
/rpt_family="MLT1"
46676. 46841
/rpt_family="MER2"
46871. 46959
/rpt_family="Alu"
complement(46983. 47265)
/rpt_family="Alu"
47306. 47395
/rpt_family="Alu"
47503. 47580
/rpt_family="Alu"
47627. 47741
repeat_region

us-09-656-668-198.oli.rge

Wed Nov 7 09:21:18 2001

```

repeat_region      /rpt_family="MER2"
complement(47741..47839)
/rpt_family="MIR"
repeat_region      49558..49671
/rpt_family="MIR"
repeat_region      50232..50508
/rpt_family="Alu"
repeat_region      complement(52633..52764)
/rpt_family="MIR"
repeat_region      complement(53356..53634)
/rpt_family="Alu"

Query Match      3.98; Score 23; DB 85; Length 127590;
Best Local Similarity 100.0%; Pred. No. 0.078;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 agaattatagcccaaatgctc 185
|||||
Db 15032 AGAATTATCAGCCCAAAATGCTC 15054

RESULT 11
AC067916 134394 bp DNA HTG 09-JUN-2000
LOCUS Homo sapiens chromosome X clone RP11-513J24 map X, *** SEQUENCING
DEFINITION IN PROGRESS ***, 32 unordered pieces.
ACCESSION AC067916
VERSION AC067916.2 GI:8389561
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 134394)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckgalter,B., Brown,A., Burgett,G.,
Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodgson,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Kleio,J., Laocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
McDrim,J., Meneus,D., Mihova,T., Miranda,C., Mienda,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivaz,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Roman,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye.W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 9, 2000 this sequence version replaced gi:7652204.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6355

```

Center clone name: 513_J_24

NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1240: contig of 1240 bp in length
1241 1340: gap of 100 bp
1341 2613: contig of 1273 bp in length
2614 2713: gap of 100 bp
2714 4159: contig of 1446 bp in length
4160 4259: gap of 100 bp
4260 5435: contig of 1176 bp in length
5436 5535: gap of 100 bp
5536 6901: contig of 1366 bp in length
6902 7001: gap of 100 bp
7002 8119: contig of 1118 bp in length
8120 8219: gap of 100 bp
8220 10213: contig of 1994 bp in length
10214 10313: gap of 100 bp
10314 12093: contig of 1780 bp in length
12094 12193: gap of 100 bp
12194 13509: contig of 1316 bp in length
13510 13609: gap of 100 bp
13610 14871: contig of 1262 bp in length
14872 14971: gap of 100 bp
14972 16966: contig of 1995 bp in length
16967 17066: gap of 100 bp
17067 18732: contig of 1666 bp in length
18733 18832: gap of 100 bp
18833 20555: contig of 1823 bp in length
20556 20755: gap of 100 bp
20756 22721: contig of 1966 bp in length
22722 22821: gap of 100 bp
22822 24991: contig of 2170 bp in length
24992 25091: gap of 100 bp
25092 25443: contig of 352 bp in length
25444 25543: gap of 100 bp
25544 28606: contig of 3063 bp in length
28607 28706: gap of 100 bp
28707 30875: contig of 2169 bp in length
30876 30975: gap of 100 bp
30976 34223: contig of 3248 bp in length
34224 34393: gap of 100 bp
34324 37911: contig of 3588 bp in length
37912 38011: gap of 100 bp
38012 42033: contig of 4022 bp in length
42034 42133: gap of 100 bp
42134 45055: contig of 2922 bp in length
45056 45155: gap of 100 bp
45156 50793: contig of 5638 bp in length
50794 50893: gap of 100 bp
50894 56638: contig of 5745 bp in length
56639 56738: gap of 100 bp
56739 62187: contig of 5449 bp in length
62188 62287: gap of 100 bp
62288 69769: contig of 7482 bp in length
69770 69869: gap of 100 bp
69870 77546: contig of 7677 bp in length
77547 77646: gap of 100 bp
77647 84723: contig of 7077 bp in length
84724 84823: gap of 100 bp
84824 92916: contig of 8093 bp in length
92917 93016: gap of 100 bp
93017 107613: contig of 14597 bp in length
107614 107713: gap of 100 bp
107714 119925: contig of 12212 bp in length
119926 120025: gap of 100 bp
120026 134394: contig of 14369 bp in length.

Wed Nov 7 09:21:18 2001

```

FEATURES
source
    Location/Qualifiers
    1..134394
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="X"
    /map="X"
    /clone="RP11-513J24"
    /clone.lib="RPC1-11 Human Male BAC"
    1..1240
    /note="assembly_fragment"
    1341..2613
    /note="assembly_fragment"
    2714..4159
    /note="assembly_fragment"
    4260..5435
    /note="assembly_fragment"
    5536..6901
    /note="assembly_fragment"
    7002..8119
    /note="assembly_fragment"
    8220..10213
    /note="assembly_fragment"
    10314..12093
    /note="assembly_fragment"
    12194..13509
    /note="assembly_fragment"
    13610..14871
    /note="assembly_fragment"
    14972..16966
    /note="assembly_fragment"
    17067..18732
    /note="assembly_fragment"
    18833..20655
    /note="assembly_fragment"
    20756..22721
    /note="assembly_fragment"
    22822..24991
    /note="assembly_fragment"
    25092..25443
    /note="assembly_fragment"
    clone_end:SP6
    vector_side:right"
    25544..28606
    /note="assembly_fragment"
    28707..30875
    /note="assembly_fragment"
    30976..34223
    /note="assembly_fragment"
    34324..37911
    /note="assembly_fragment"
    38012..42033
    /note="assembly_fragment"
    42134..445055
    /note="assembly_fragment"
    clone_end:T7
    vector_side:right"
    45156..50793
    /note="assembly_fragment"
    50894..55638
    /note="assembly_fragment"
    56739..62187
    /note="assembly_fragment"
    62288..69769
    /note="assembly_fragment"
    69870..77546
    /note="assembly_fragment"
    77647..84723
    /note="assembly_fragment"
    84824..92916
    /note="assembly_fragment"
    93017..107613
    /note="assembly_fragment"
    107714..119925
    /note="assembly_fragment"
    120026..134394
    /note="assembly_fragment"
    25781 c 26014 g 40164 t 3107 others
    BASE COUNT 39328 a 25781 c 26014 g 40164 t 3107 others
    Query Match 3.8%; Score 23; DB 73; Length 134394;
    Best Local Similarity 100.0%; Pred. No. 0.078;
    Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
    QY 163 agaatattctagcccaaatgtc 185
    |||
    Db 77870 AGAATTATCTAGCCCAATGTC 77848

RESULT 12
AC011573 151357 bp DNA HTG 12-MAR-2000
LOCUS Homo sapiens clone RP11-12G12, WORKING DRAFT SEQUENCE, 28 unordered
DEFINITION pieces.
ACCESSION AC011573
VERSION AC011573.3 GI:7230070
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 151357)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-12G12
2 (bases 1 to 151357)
Unpublished
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Deatellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferrelira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Horton,L.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Klein,J.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Marquis,N.,
Lehoczky,J., Lieu,C., Locke,K., McKernan,K., McLaughlin,J., Meldrim,J.,
McEwan,P., McGurk,A., McKernan,C.H., O'Connor,T., O'Donnell,P.,
Morrow,J., Naylor,J., Norman,C.H., Roy,A., Santos,R., Severy,P.,
Peterson,K., Pollara,V., Riley,R., Riley,C., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (07-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:6249718.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3265
Center clone name: 12.G.12
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 101368 bases at least Q40
Consensus quality: 127052 bases at least Q30
Consensus quality: 139101 bases at least Q20
Insert size: 147000; agarose-fp
Insert size: 148657; sum-of-contigs
Quality coverage: 3.5 in Q20 bases; agarose-fp
Quality coverage: 3.5 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently

```

us-09-656-668-198.oli.rge

Wed Nov 7 09:21:18 2001

* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1505: contig of 1505 bp in length
* 1506 1605: gap of 100 bp
* 1606 2995: contig of 1390 bp in length
* 2996 3095: gap of 100 bp
* 3096 4668: contig of 1573 bp in length
* 4669 4768: gap of 100 bp
* 4769 6111: contig of 1343 bp in length
* 6112 6211: gap of 100 bp
* 6212 7559: contig of 1348 bp in length
* 7560 7659: gap of 100 bp
* 7660 8904: contig of 1245 bp in length
* 8905 9004: gap of 100 bp
* 9005 10441: contig of 1437 bp in length
* 10442 10541: gap of 100 bp
* 10542 12992: contig of 2451 bp in length
* 12993 13092: gap of 100 bp
* 13093 16001: contig of 2909 bp in length
* 16002 16101: gap of 100 bp
* 16102 18808: contig of 2707 bp in length
* 18809 18908: gap of 100 bp
* 18909 22389: contig of 3481 bp in length
* 22390 22489: gap of 100 bp
* 22490 25729: contig of 3240 bp in length
* 25730 25929: gap of 100 bp
* 25930 28948: contig of 3119 bp in length
* 28949 29048: gap of 100 bp
* 29049 32582: contig of 3534 bp in length
* 32583 32682: gap of 100 bp
* 32683 36104: contig of 3422 bp in length
* 36105 36204: gap of 100 bp
* 36205 39825: contig of 3621 bp in length
* 39826 39925: gap of 100 bp
* 39926 45184: contig of 5259 bp in length
* 45185 45284: gap of 100 bp
* 45285 50916: contig of 5632 bp in length
* 50917 51016: gap of 100 bp
* 51017 56385: contig of 5369 bp in length
* 56386 56485: gap of 100 bp
* 56486 63439: contig of 6954 bp in length
* 63440 63539: gap of 100 bp
* 63540 70246: contig of 6707 bp in length
* 70247 70346: gap of 100 bp
* 70347 76319: contig of 5973 bp in length
* 76320 76419: gap of 100 bp
* 76420 85131: contig of 8712 bp in length
* 85132 85231: gap of 100 bp
* 85232 91974: contig of 6743 bp in length
* 91975 92074: gap of 100 bp
* 92075 102742: contig of 10668 bp in length
* 102743 102842: gap of 100 bp
* 102843 110539: contig of 7697 bp in length
* 110540 110639: gap of 100 bp
* 110640 119486: contig of 8847 bp in length
* 119487 119586: gap of 100 bp
* 119587 151357: contig of 31771 bp in length.

FEATURES
source
1. .151357
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-12G12"
/clone_lib="RPC1-11 Human Male BAC"
1. .1505
/note="assembly_fragment"
1606. .2995
/note="assembly_fragment"
3096 4668
/note="assembly_fragment"

misc_feature /note="assembly_fragment"
4769. .6111
/note="assembly_fragment"
6212. .7559
/note="assembly_fragment"
7660. .8904
/note="assembly_fragment"
9005. .10441
/note="assembly_fragment"
10542. .12992
/note="assembly_fragment"
13093. .16001
/note="assembly_fragment"
16102. .18808
/note="assembly_fragment"
18909. .22389
/note="assembly_fragment"
22490. .25729
/note="assembly_fragment"
25830. .28948
/note="assembly_fragment"
29049. .32582
/note="assembly_fragment"
32683. .36104
/note="assembly_fragment"
36205. .39825
/note="assembly_fragment"
39926. .45184
/note="assembly_fragment"
45285. .50916
/note="assembly_fragment"
51017. .56385
/note="assembly_fragment"
56486. .63439
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
63540. .70246
/note="assembly_fragment"
70347. .76319
/note="assembly_fragment"
76420. .85131
/note="assembly_fragment"
85232. .91974
/note="assembly_fragment"
clone_end:R7
vector_side:left"
92075. .102742
/note="assembly_fragment"
102843. .110539
/note="assembly_fragment"
110640. .119486
/note="assembly_fragment"
119587. .151357
/note="assembly_fragment"
BASE COUNT 40770 a 33565 c 33412 g 40768 t 2842 others
ORIGIN
Query Match 3.8%; Score 23; DB 62; Length 151357;
Best Local Similarity 100.0%; Pred. No. 0.078;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 56 ttttgattgtcacacctgggtgg 78
|||||
Db 56633 TTTTGATTGTACACCTGGGTGG 56655
RESULT 13
AC034161 157402 bp DNA HTG 01-SEP-2000
LOCUS Homo sapiens chromosome 11 clone RP11-397J5, WORKING DRAFT
DEFINITION SEQUENCE, 18 unordered pieces.

us-09-656-668-198.oli.rge

Wed Nov 7 09:21:18 2001

AC034161
AC034161.3 GI:9958241
HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 157402)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 157402)
Waterston,R.H.
Direct Submission
Submitted (04-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Sep 1, 2000 this sequence version replaced gi:7523965.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0397J05
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 149716 bases at least Q40
Consensus quality: 152398 bases at least Q30
Consensus quality: 153456 bases at least Q20
Insert size: 191000; agarose-fp
Insert size: 155702; sum-of-contigs
Quality coverage: 3.43 in Q20 bases; agarose-fp
Quality coverage: 4.25 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 5077: contig of 5077 bp in length
* 5078 5177: gap of unknown length
* 5178 11207: contig of 6030 bp in length
* 11208 11307: gap of unknown length
* 11308 18089: contig of 6782 bp in length
* 18090 18189: gap of unknown length
* 18190 24820: contig of 6631 bp in length
* 24821 24820: gap of unknown length
* 24821 32091: contig of 7171 bp in length
* 32092 32191: gap of unknown length
* 32192 45992: contig of 13801 bp in length
* 45993 46092: gap of unknown length
* 46093 52957: contig of 6865 bp in length
* 52958 53058: gap of unknown length
* 53059 61361: contig of 8303 bp in length
* 61362 61460: gap of unknown length
* 61461 69613: contig of 8153 bp in length
* 69614 80408: contig of 10695 bp in length
* 80409 80508: gap of unknown length
* 80509 91848: contig of 11340 bp in length
* 91849 91948: gap of unknown length
* 91949 115447: contig of 23499 bp in length
* 115448 115547: gap of unknown length
* 115548 142318: contig of 26771 bp in length
* 142319 142319: gap of unknown length

142419 145665: contig of 3247 bp in length
* 145666 145765: gap of unknown length
* 145766 148715: contig of 2950 bp in length
* 148716 148815: gap of unknown length
* 148816 151082: contig of 2267 bp in length
* 151083 151182: gap of unknown length
* 151183 153839: contig of 2657 bp in length
* 153840 153939: gap of unknown length
* 153940 157402: contig of 3463 bp in length.
Location/Qualifiers
1. .157402
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP11-397J5"
1. .5077
/note="assembly_name:Contig10"
5178. .11207
/note="assembly_name:Contig11"
11308. .18089
/note="assembly_name:Contig12"
18190. .24820
/note="assembly_name:Contig13"
24921. .32091
/note="assembly_name:Contig14"
clone_end.SP6
vector_side:right"
32192. .45992
/note="assembly_name:Contig19"
clone_end.T7
vector_side:right"
46093. .52957
/note="assembly_name:Contig15"
53058. .61360
/note="assembly_name:Contig16"
61461. .69613
/note="assembly_name:Contig17"
69714. .80408
/note="assembly_name:Contig18"
80509. .91848
/note="assembly_name:Contig20"
91949. .115447
/note="assembly_name:Contig21"
115548. .142318
/note="assembly_name:Contig22"
142419. .145665
/note="assembly_name:Contig5"
145766. .148715
/note="assembly_name:Contig6"
148816. .151082
/note="assembly_name:Contig7"
151183. .153839
/note="assembly_name:Contig8"
153940. .157402
/note="assembly_name:Contig9"
BASE COUNT 49815 a 29500 c 29459 g 46922 t 1706 others
ORIGIN
Query Match 3.8%; Score 23; DB 71; Length 157402;
Best Local Similarity 100.0%; Pred. No. 0.078;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 163 aqaattatctagccccaatgtc 185
|||||
Db 13008 AGAATTATCTAGCCCAATGTC 13030
RESULT 14
AC021862 159723 bp DNA HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 11 clone RP11-674E24, WORKING DRAFT
DEFINITION SEQUENCE, 18 unordered pieces.

us-09-656-668-198.oli.rge

Wed Nov 7 09:21:18 2001

AC021862
 VERSION AC021862.4 GI:8570373
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 159723)
 Waterston,R.H.
 The sequence of Homo sapiens clone
 2 (bases 1 to 159723)
 Waterston,R.H.
 Direct Submission
 Submitted (20-JAN-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 On Jun 17, 2000 this sequence version replaced gi:7235315.

COMMENT

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 ----- Project Information -----
 Center project name: H_NH0674E24
 ----- Summary Statistics -----
 Sequencing vector: M13, 76%
 Sequencing vector: plasmid; 24%
 Chemistry: Dye-primer ET; 76% of reads
 Chemistry: Dye-terminator Big Dye; 24% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 146571 bases at least Q40
 Consensus quality: 150482 bases at least Q30
 Consensus quality: 152816 bases at least Q20
 Insert size: 167000; agarose-fp
 Insert coverage: 3.44 in Q20 bases; sum-of-contigs
 Quality coverage: 3.71 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 18 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1195: contig of 1195 bp in length
 * 1196 1295: gap of unknown length
 * 1296 3879: contig of 2584 bp in length
 * 3880 3979: gap of unknown length
 * 3980 6597: contig of 2618 bp in length
 * 6598 10040: gap of unknown length
 * 10041 10140: contig of 3343 bp in length
 * 10141 12973: contig of 2833 bp in length
 * 12974 13074: gap of unknown length
 * 13074 15987: contig of 2914 bp in length
 * 15988 16087: gap of unknown length
 * 16088 19014: contig of 2927 bp in length
 * 19015 19114: gap of unknown length
 * 19115 22164: contig of 3050 bp in length
 * 22165 22664: gap of unknown length
 * 22665 26798: contig of 4534 bp in length
 * 26799 26898: gap of unknown length
 * 26899 34198: contig of 7300 bp in length
 * 34199 34298: gap of unknown length
 * 34299 42840: contig of 8542 bp in length
 * 42841 42841: gap of unknown length
 * 42841 48833: contig of 5893 bp in length
 * 48834 59689: gap of unknown length
 * 59690 59789: contig of 10756 bp in length
 * 59789 59789: gap of unknown length

59790 72021: contig of 12232 bp in length
 * 72022 72121: gap of unknown length
 * 72122 87123: contig of 15002 bp in length
 * 87124 87223: gap of unknown length
 * 87224 106312: contig of 19089 bp in length
 * 106313 106412: gap of unknown length
 * 106413 133985: contig of 27573 bp in length
 * 133986 134086: gap of unknown length
 * 134086 159723: contig of 25638 bp in length.
 Location/Qualifiers
 1. 159723
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /clone="RP11-674E24"
 1. 1195
 /note="assembly_name:Contig7"
 1296. 3879
 /note="assembly_name:Contig8"
 3980. 6597
 /note="assembly_name:Contig9"
 6698. 10040
 /note="assembly_name:Contig10"
 clone_end:SP6
 vector_side:left
 10141. 12973
 /note="assembly_name:Contig11"
 13074. 15987
 /note="assembly_name:Contig12"
 16088. 19014
 /note="assembly_name:Contig13"
 19115. 22164
 /note="assembly_name:Contig14"
 22265. 26798
 /note="assembly_name:Contig15"
 26899. 34198
 /note="assembly_name:Contig16"
 34299. 42840
 /note="assembly_name:Contig17"
 42941. 48833
 /note="assembly_name:Contig18"
 48934. 59689
 /note="assembly_name:Contig19"
 59790. 72021
 /note="assembly_name:Contig20"
 72122. 87123
 /note="assembly_name:Contig21"
 87224. 106312
 /note="assembly_name:Contig22"
 106413. 133985
 /note="assembly_name:Contig23"
 134086. 159723
 /note="assembly_name:Contig24"
 1728 others
 BASE COUNT 48651 a 29918 c 29977 g 49449 t

Query Match 3.8% Score 23; DB 66; Length 159723;
 Best Local Similarity 100.0%; Pred. No. 0.078;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 163 agaattatctagcccaaatgtc 185
 |||||
 Db 3629 AGAATTATCTAGCCCAATGTCTC 3651
 RESULT 15
 AC078953 171589 bp DNA HTG 07-JAN-2001
 LOCUS Homo sapiens chromosome 3q clone RP11-370M22, *** SEQUENCING IN
 DEFINITION PROGRESS ***, 61 unordered pieces.
 ACCESSION AC078953
 VERSION AC078953.12 GI:12039127

Wed Nov 7 09:21:18 2001

HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens

KEYWORDS SOURCE ORGANISM

REFERENCE
AUTHORS

1 (bases 1 to 171589)
Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Blum,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Bowie,S., Brieve,M., Brown,E., Byrd,N.C., Carron,T.F.,
Burch,P., Burkett,C., Burrell,K.L., Chavez,D., Chen,G., Chen,R.,
Carter,M., Cavazos,S.R., Chacko,J., Chavaz,D., Cleveland,C.D., Cox,C.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Davis,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Correll,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Hernandez,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hughes,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Jolivet,S.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseg,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Ogih,M., Okwundu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Picketts,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojebokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Umani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williams,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 171589)
Worley,K.C.
Direct Submission
Submitted (13-AUG-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jan 5, 2001 this sequence version replaced gi:11995537.

Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

Project information

Center project name: HBS2
Center clone name: RP11-370M22

Summary Statistics

Sequencing vector: M13, L08821
Chemistry: Dye-terminator Big Dye: 48% of reads
Assembly: Dye-terminator Big Dye: 48% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 135345 bases at least Q40
Consensus quality: 149003 bases at least Q30
Consensus quality: 157760 bases at least Q20
Estimated coverage: 151105: sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 2.5x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 61 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 9607: contig of 9607 bp in length
9608 9707: gap of unknown length
15420 15420: contig of 5713 bp in length
15520 15520: gap of unknown length
21934 21934: contig of 6414 bp in length
22034 22034: gap of unknown length
28096 28096: contig of 6062 bp in length
28097 28097: gap of unknown length
33873 33873: contig of 5677 bp in length
33874 33874: gap of unknown length
39416 39416: contig of 5443 bp in length
39417 39417: gap of unknown length
44350 44350: contig of 4833 bp in length
44351 44351: gap of unknown length
44352 44352: contig of 4215 bp in length
48664 48664: gap of unknown length
48665 48665: contig of 4910 bp in length
48765 48765: gap of unknown length
53675 53675: gap of unknown length
58552 58552: contig of 4877 bp in length
58553 58553: gap of unknown length
63000 63000: contig of 4249 bp in length
63001 63001: gap of unknown length
66683 66683: contig of 3583 bp in length
66783 66783: gap of unknown length
69705 69705: contig of 2922 bp in length
69805 69805: gap of unknown length
73159 73159: contig of 3354 bp in length
73259 73259: gap of unknown length
76844 76844: contig of 3585 bp in length
76944 76944: gap of unknown length
80240 80240: contig of 3296 bp in length
80340 80340: gap of unknown length
83170 83170: contig of 2830 bp in length
83270 83270: gap of unknown length
85770 85770: contig of 2500 bp in length
85870 85870: gap of unknown length
88260 88260: contig of 2390 bp in length
88360 88360: gap of unknown length
91251 91251: contig of 2891 bp in length
91252 91252: gap of unknown length
91351 91351: contig of 3787 bp in length
95139 95139: gap of unknown length
95238 95238: gap of unknown length
97514 97514: contig of 2276 bp in length
97614 97614: gap of unknown length
99783 99783: contig of 2169 bp in length
99883 99883: gap of unknown length
102499 102499: contig of 2616 bp in length
102599 102599: gap of unknown length
105018 105018: contig of 2419 bp in length
105118 105118: gap of unknown length
107726 107726: contig of 2608 bp in length
107826 107826: gap of unknown length
111312 111312: contig of 3486 bp in length
111412 111412: gap of unknown length
114820 114820: contig of 3408 bp in length
114920 114920: gap of unknown length
117190 117190: contig of 2270 bp in length
117290 117290: gap of unknown length
119277 119277: contig of 1987 bp in length
119377 119377: gap of unknown length
120960 120960: contig of 1583 bp in length
121060 121060: gap of unknown length

Wed Nov 7 09:21:18 2001

*	121061	122181:	contig of 1121 bp in length
*	122182	122281:	gap of unknown length
*	122282	124038:	contig of 1757 bp in length
*	124039	124138:	gap of unknown length
*	124139	125666:	contig of 1528 bp in length
*	125667	125766:	gap of unknown length
*	125767	127725:	contig of 1959 bp in length
*	127726	127825:	gap of unknown length
*	127826	130376:	contig of 2551 bp in length
*	130377	130476:	gap of unknown length
*	130477	132820:	contig of 2344 bp in length
*	132821	132920:	gap of unknown length
*	132921	135499:	contig of 2579 bp in length
*	135500	135599:	gap of unknown length
*	135600	137202:	contig of 1603 bp in length
*	137203	137302:	gap of unknown length
*	137303	139112:	contig of 1810 bp in length
*	139113	139212:	gap of unknown length
*	139213	140413:	contig of 1201 bp in length
*	140414	140513:	gap of unknown length
*	140514	142609:	contig of 2036 bp in length
*	142610	142709:	gap of unknown length
*	142710	144637:	contig of 1928 bp in length
*	144638	144737:	gap of unknown length
*	144738	146091:	contig of 1354 bp in length
*	146092	146191:	gap of unknown length
*	146192	147858:	contig of 1667 bp in length
*	147859	147958:	gap of unknown length
*	147959	149631:	contig of 1673 bp in length
*	149632	149731:	gap of unknown length
*	149732	150740:	contig of 1009 bp in length
*	150741	150840:	gap of unknown length
*	150841	152348:	contig of 1508 bp in length
*	152349	152448:	gap of unknown length
*	152449	153976:	contig of 1528 bp in length
*	153977	155077:	gap of unknown length
*	155077	155379:	contig of 1303 bp in length
*	155380	155479:	gap of unknown length
*	155480	157688:	contig of 2209 bp in length
*	157689	157788:	gap of unknown length
*	157789	159085:	contig of 1297 bp in length
*	159086	159185:	gap of unknown length
*	159186	160288:	contig of 1103 bp in length
*	160289	160388:	gap of unknown length
*	160389	162315:	contig of 1927 bp in length
*	162316	162415:	gap of unknown length
*	162416	165543:	contig of 1128 bp in length
*	165544	163643:	gap of unknown length
*	163644	164773:	contig of 1130 bp in length

Query Match	3.8%	Score 23;	DB 75;	Length 1/1589;
Best Local Similarity	100.0%;	Pred. No. 0.079;		
Mismatches 23;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;

QY 163 agaattatctagccccaaatgtc 185
|||||
Db 1643 AGAATTATCTAGCCCCAAATGTC 1665

Search completed: November 5, 2001, 23:44:59
Job time: 2829 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 5, 2001, 22:33:25 ; Search time 126.12 Seconds
(without alignments)
3017.034 Million cell updates/sec

Title: US-09-656-668-198
Perfect score: 606
Sequence: 1 tgagtttgcccttaccctcc.....aagctgttcttgcgcac 606

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 730101 seqs, 313950809 residues
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	606	100.0	606	22	AAF95006 Human ovarian can
2	82.4	13.6	80240	20	AAV83940 NC-contig derived
3	82.4	13.6	80595	20	AAV83939 HC-contig derived
4	73	12.0	2099	21	AAA27046 Human cell surface
5	73	12.0	2116	21	AAAC59845 Human secreted pro
6	71	11.7	2061	20	AAZ33929 Human PRO329 nucle
7	71	11.7	2061	21	AAZ78471 Human PRO329 (UNC2
8	68.2	11.3	208	21	AAAF15763 Human prostate can
9	67.4	11.1	1829	21	AAAC69639 Human secreted pro
10	66.2	10.9	11950	20	AAAX90201 Human yes1 gene.
11	65.8	10.9	3280	21	AAA26367 Human secreted pro

12	65.8	10.9	3300	21	AAA26449 Human secreted pro
13	64.8	10.7	7505	20	AAV83949 Bacterial artifi
14	64.6	10.7	1583	21	AAAC79962 Human secreted pro
15	63.4	10.5	2020	21	AAAC93321 Human neuroblastom
16	61.2	10.1	1119	22	AAAF97864 Human kinesin-like
17	61.2	10.1	121162	21	AAAC66548 Human KIK-L2 gene.
18	60.6	10.0	11570	21	AAAC95905 Sequence of a cDNA
19	60.4	10.0	6314	14	AAQ40341 Nucleotide sequenc
20	60	9.9	50000	21	AAAC64139 Homo sapiens 20q13
21	56.2	9.3	10282	19	AAV09023 Retinoblastoma bin
22	56.2	9.3	162450	21	AAZ86967 Human biallelic po
23	56	9.2	251	19	AAAX10977 Human secreted pro
24	54.8	9.0	1920	21	AAAC59071 Human secreted pro
25	54.8	9.0	1921	22	AAAF32739 Human secreted pro
26	54.8	9.0	2499	21	AAAC80607 Human secreted pro
27	54.8	9.0	2522	21	AAAC80618 Lung cancer associ
28	54.8	9.0	2547	21	AAAF18213 Human ORFX ORF2097
29	54.6	9.0	6439	21	AAAF76542 Human flavin-conta
30	54.2	8.9	125910	21	AAAC64370 Human kidney amino
31	53.8	8.9	25464	19	AAV57274 Human secreted pro
32	53.2	8.8	49998	20	AAAC23518 Novel human polynu
33	53	8.7	354	21	AAAC22185 Human secreted pro
34	52.8	8.7	381	22	AAAF66536 Human secreted pro
35	52.8	8.7	476	21	AAAC20444 Human secreted pro
36	52.6	8.7	693	20	AAAC16692 Human secreted pro
37	52.6	8.7	155	21	AAAC97944 Human secreted pro
38	52	8.6	438	20	AAAC98795 Human pancreatic c
39	52	8.6	2423	21	AAAC98795 Oligonucleotide D1
40	51.6	8.5	936	22	AAAF58252 Oligonucleotide D1
41	51.6	8.5	936	22	AAAF58257 Oligonucleotide D2
42	51.6	8.5	936	22	AAAF58259 Oligonucleotide D2
43	51.6	8.5	936	22	AAAF58262 Oligonucleotide D1
44	51.6	8.5	936	22	AAAF58265 Oligonucleotide D1
45	51.6	8.5	938	22	AAAF58255 Oligonucleotide D1

ALIGNMENTS

RESULT 1
AAF95006
ID AAF95006 standard; DNA; 606 BP.
XX
AC AAF95006;
XX
DT 23-MAY-2001 (first entry)
XX
DE Human ovarian cancer associated coding sequence SEQ ID NO: 198.
XX
KW Human, ovarian cancer; vaccine; gene therapy; carcinoma; ds.
XX
OS Homo sapiens.
XX
PN WO200118046-A2.
XX
PD 15-MAR-2001.
XX
PF 08-SEP-2000; 2000WO-US24827.
XX
PR 10-SEP-1999; 99US-0394374.
PR 01-MAY-2000; 2000US-0561778.
PR 15-AUG-2000; 2000US-0640173.
PR 07-SEP-2000; 2000US-0656668.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Stolk JA;
XX
DR WFI; 2001-211395/21.
XX
PT Isolated polypeptides associated with ovarian carcinomas, and the
PT nucleic acids that encode them, useful for the prevention diagnosis and
PT treatment of ovarian cancers -

Wed Nov 7 09:21:20 2001

XX PS Claim 18; Page 189; 189pp; English.

XX CC The present invention provides a number of coding sequences and proteins,

CC CC the over-expression of which is associated with ovarian carcinoma/cancer.

CC CC These can be used in the diagnosis, treatment and prevention of ovarian

CC CC cancer, optionally by gene therapy or in the form of a vaccine. The

CC CC present sequence is an example of one of these sequences.

XX SQ Sequence 606 BP; 175 A; 147 C; 123 G; 161 T; 0 other;

Query Match 100.0%; Score 606; DB 22; Length 606;

Best Local Similarity 100.0%; Pred. No. 7e-187;

Matches 606; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgaattgccccttaccctccatccagtggaattttgcaattcttaagagcgtgtttg 60

Db 1 tgaattgccccttaccctccatccagtggaattttgcaattcttaagagcgtgtttg 60

QY 61 attgtcacacctgggtgggaacatgctactgcatctaatgcatagaggcggaatgc 120

Db 61 attgtcacacctgggtgggaacatgctactgcatctaatgcatagaggcggaatgc 120

QY 121 tgcataacattttcaacgcacagacagagagcccccacaaagagaattatctagcccaa 180

Db 121 tgcataacattttcaacgcacagacagagagcccccacaaagagaattatctagcccaa 180

QY 181 atgtcataaacactgctgttgagaacacctccagcagatcttactgggttcctatagta 240

Db 181 atgtcataaacactgctgttgagaacacctccagcagatcttactgggttcctatagta 240

QY 241 agcttgcctttgttgccttctgtagatatataaaataaagacactgcccagtcctcc 300

Db 241 agcttgcctttgttgccttctgtagatatataaaataaagacactgcccagtcctcc 300

QY 301 ctcaacgtcccgagccagggctcaaggcaattcccaataacagtagaataaactaaata 360

Db 301 ctcaacgtcccgagccagggctcaaggcaattcccaataacagtagaataaactaaata 360

QY 361 ttgatttcaaaatctcagcaactagagaatgaccaaaccctcctggttggcctggactg 420

Db 361 ttgatttcaaaatctcagcaactagagaatgaccaaaccctcctggttggcctggactg 420

QY 421 tccagttttagcattgaaagtcttcaggttccaggaaagccctcagcctggctgtg 480

Db 421 tccagttttagcattgaaagtcttcaggttccaggaaagccctcagcctggctgtg 480

QY 481 tcaacctagcagctgaggactcttcaatacagaattagtttgcactggagatgaa 540

Db 481 tcaacctagcagctgaggactcttcaatacagaattagtttgcactggagatgaa 540

QY 541 tatactttaattgttaacatgtaaacatctataaacatctactgaagcctgtttgt 600

Db 541 tatactttaattgttaacatgtaaacatctataaacatctactgaagcctgtttgt 600

QY 601 ctgcaac 606

Db 601 ctgcaac 606

RESULT 2

AAV83940

ID AAV83940 standard; DNA: 80240 BP.

XX AC AAV83940;

XX DT 03-MAR-1999 (first entry)

XX DE NC-contig derived from mardel(10) on chromosome 10q25.2.

XX X' Yeast artificial chromosome; YAC; probe: eukaryotic chromosome;

KW neocntromere; replication; extra-chromosomal element; segregation;

KW cell division; artificial chromosome; gene therapy; mardel(10);

KW human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.

XX Homo sapiens.

XX WO9851790-A1.

PN 19-NOV-1998.

PD 13-MAY-1998;

XX 98WO-AU00352.

PF 26-AUG-1997;

XX 97AU-0008791.

PR 13-MAY-1997;

PR 97AU-0006784.

XX (AMRA-) AMRAD OPERATIONS PTY LTD.

PA Cancellia MR, Choo K, Du Sart D;

PI WPI; 1999-009773/01.

XX New isolated nucleic acid comprising neocentromere sequences from

DR eukaryotic chromosome - used to produce replicable, segregating

PT artificial chromosomes that can carry large amounts of DNA for gene

PT therapy

XX Claim 9; Fig 16A; 540pp; English.

PS The present sequence represents the NC-contig derived from a mutated

XX human chromosome 10, 10q25.2 region. The sequence contains

CC an unusual chromosomal marker referred to as mardel(10). The

CC mardel(10) marker is mitotically stable and contains a functional

CC neocentromere at a location regarded as non-centromeric. This

CC neocentromere maps to q25.2 on chromosome 10. The specification describes

CC nucleic acid sequences derived from a eukaryotic chromosome, including a

CC neocentromere or its functional derivative or hybrid, that are able, in

CC a compatible cell, of replicating, acting as extra-chromosomal element

CC and segregating during cell division. The sequences can be used to

CC construct artificial chromosomes for use in gene therapy comprising a

CC replicable, segregating nucleic acid that confers a specific phenotype

CC on cells. Human artificial chromosomes can propagate in human cells and

CC carry large amounts of DNA (e.g. therapeutic genes), and, being

CC extra-chromosomal, they are not mutagenic. The artificial chromosomes

CC are also useful for generation of transgenic plants and animals, in

CC production of cytokines, receptors and growth factors, or to increase

CC expression of proteins and to make diagnostic reagents, e.g. for

CC the copy number of a gene in a cell. The constructs may also be

CC used for functional and structural analysis of chromosomes.

XX SQ Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T; 8 other;

Query Match 13.6%; Score 82.4; DB 20; Length 80240;

Best Local Similarity 63.8%; Pred. No. 1.2e-15;

Matches 125; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 72 tgggtgggaacatgctactgcatctaatgcatagaggcgagtaatgctgctaacaac 131

Db 66530 tgaagtgggtgggtgctactgcatctagtgggtggagacagagatgctttaaacatc 66589

QY 132 tttaacgcacagacagagagcccccacaaagagaattatctagccccaaatgtccataac 191

Db 66590 ccgcaaacagacag 66649

QY 192 actgctgttgagaacacctaccgcagagagatcttactgggttcattaggaagctgcttt 251

Db 66650 gccaaagtggagaacacctcattctagcttcttcttcttcttcttcttcttcttctt 66709

QY 252 gttctggcttctgtag 267

Db 66710 gttcttcagcattag 66725

RESULT 3

QY	192	actgctgttagaataaacctaccgagatcttacttctggttcattaggttaagcttgcttt	251
Db	66908	gcacaagttgagaaacctcaattctagcttctcttccctctctcttcttaatacaactgtt	66967
QY	252	gttcctggcttctgtag	267
Db	66968	gtctcttcagcattag	66983
RESULT	4		
AA227046			
ID	AA227046	standard; cDNA; 2099 BP.	
AC	AA227046;		
CC			
DT	22-AUG-2000	(first entry)	
DE	Human cell surface receptor protein cDNA sequence #3.		
KW	Human; HCSR; cytotatic; antiarthritic; antirheumatic; antihemmatic; antiasthmatic;		
KW	Immunosuppressive; antiarteriosclerotic; antibacterial; antiparasitic;		
KW	neuroprotective; nootropic; anticonvulsant; cancer; leukaemia;		
KW	melanoma; rheumatoid arthritis; asthma; atherosclerosis; akathesia;		
KW	Alzheimer's diseases; multiple sclerosis; epilepsy; ss.		
OS	Homo sapiens.		
XX			
XX	Key	Location/Qualifiers	
FH	CDS	43..1119	
FT		/*tag= a	
FT		/product= "HCSR-3"	
XX			
PN	W0200028032-A2.		
XX			
PD	18-MAY-2000.		
XX			
PF	12-NOV-1999;	99W0-US26742.	
XX			
PR	12-NOV-1998;	98US-0191280.	
PR	07-DEC-1998;	98US-0206647.	
PR	08-MAR-1999;	99US-0123404.	
XX			
PA	(INCY-) INCYTE PHARM INC.		
XX			
PI	Tang YT, Corley NC, Guedler KJ, Yue H, Baughn MR, Lal P;		
PI	Hillman JL, Bandman O, Azimzai Y, Au-Young J;		
XX			
DR	WPI; 2000-376546/32.		
DR	P-PSDB; AAY94336.		
XX			
PT	New human cell surface receptor protein and polynucleotide useful for		
PT	diagnosis, prevention and treatment of cancer, immune disorders,		
PT	infection and neuronal disorders -		
XX			
PS	Claim 9; Page 89-90; 97pp; English.		
XX			
CC	The present sequence encodes a novel human cell surface receptor		
CC	protein (HCSR) designated HCSR-3. The nucleotide sequence was		
CC	identified in Incyte Clone 53181 from the cDNA library LMOONNOT02,		
CC	which was made from RNA isolated from lymph node tissue. A number		
CC	of Incyte Clones were used to assemble the consensus sequence.		
CC	BLAST analysis showed that the sequence is homologous to IgG Fc		
CC	receptor 9593604. HCSR and its antagonist are useful for preventing		
CC	treating disorders associated with decreased or increased expression		
CC	activity of HCSR. Such disorders include cancers such as leukaemia		
CC	melanoma, immune disorders such as rheumatoid arthritis, asthma and		
CC	atherosclerosis, bacterial and parasitic infections and neuronal		
CC	disorders such as akathesia, Alzheimer's disease, multiple sclerosis		
CC	epilepsy. Polynucleotides encoding HCRPs may be used as hybridisa		
CC	probes to diagnose these conditions. Anti-HCSR antibodies may be u		
CC	as antagonists, as a targeting or delivery mechanism for bringing		
CC	pharmaceutical agents into contact with cells or tissues expressing		

	AAV83939 standard; DNA; 80595 BP.
ID	AAV83939
XX AC	(first entry)
XX AC	
XX AC	
XX DT	
XX DE	HC-contig derived from normal human chromosome 10q25.2 region.
DE	
XX OS	Homo sapiens.
KW KW	Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW KW	neocentromere; replication; extra-chromosomal element; segregation;
KW KW	cell division; artificial chromosome; gene therapy; mardel(10);
KW KW	human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX PN	
XX PD	
XX PF	
XX PR	
XX PA	(AMRA-) AMRAD OPERATIONS PVT LTD.
XX PI	Cancilla MR, Choo K, Du Sart D;
XX PT	WPI; 1999-009773/01.
PT	New isolated nucleic acid comprising neocentromere sequences from
PT	eukaryotic chromosome - used to produce replicable, segregating
PT	artificial chromosomes that can carry large amounts of DNA for gene
PT	therapy
PS	Claim 8; Fig 6; 54Opp; English.
CC CC	The present sequence represents the HC-contig derived from normal human
CC CC	chromosome 10, 10q25.2 region. This region can be naturally mutated to
CC CC	produce an unusual chromosomal marker referred to as mardel(10). The
CC CC	mardel(10) marker is mitotically stable and contains a functional
CC CC	neocentromere at a location regarded as non-centromeric. This
CC CC	nucleic acid sequences derived from a eukaryotic chromosome, including a
CC CC	comparable cell, of replicating, acting as extra-chromosomal element
CC CC	and segregating during cell division. The sequences can be used to
CC CC	construct artificial chromosomes for use in gene therapy comprising a
CC CC	replicable, segregating nucleic acid that confers in human cells and
CC CC	on cells. Human artifacts of DNA (e.g. therapeutic genes), and, being
CC CC	extra-chromosomal, they are not mutagenic. The artificial chromosomes
CC CC	are also useful for generation of transgenic plants and animals, in
CC CC	production of proteins and to make diagnostic reagents, e.g. for
CC CC	expression of cytokines, receptors and growth factors, or to increase
CC CC	the copy number of a gene in a cell. The constructs may also be
CC CC	used for functional and structural analysis of chromosomes.
XX SQ	Sequence 80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 other;
SQ	
	Query Match 13.6%; Score 82.4; DB 20; Length 80595;
	Best Local Similarity 63.8%; Pred. No. 1.2e-15;
Matches	Conservative 0; Mismatches 71; Indels 0; Gaps
QY	72 tgggtgagggaacatgctactggcatcctaagtgcataagaggcgagtaattgctgtaaaccatc 131
DG	66788 tgagtgggtgggttgccttgccatctagtgggtgggacagagaatttcgccaaatgccataaac 66947
QY	132 ttccaacycacaggacagagccccacaagaagaattatctagcccgaatgtccataac 191
NB	66848 ccqaagacacaggacagtcctcccgaccaaaaagaattatctggycccaaatcataagt 66907

sequences AAB34746-B34771 represent fragments of the proteins. Probes for the DNA sequences are represented by sequences AAC59847-599596. The proteins exhibit neuroprotective, dermatological, immunosuppressive, antiinflammatory, antianaemic, nootropic, antiparkinsonian, cerebroprotective, haemostatic, vulnerary, cytoplastic, antipsoriatic, antibacterial, virucide, and fungicide activity. The proteins and nucleotide sequences are useful as nutritional sources or supplements and in research. The proteins are useful for treating immune deficiency and disorders, which may be genetic or resulting from infections, autoimmune disorders such as multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, and for treating myeloid or lymphoid cell deficiencies such as anaemias by regulating haematopoiesis. The proteins are also useful in compositions for bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, for wound healing, tissue repair and replacement and in the treatment of wounds, incisions and ulcers. Other uses include in the treatment of central and peripheral nervous system and neuropathies such as Alzheimer's and Parkinson's diseases and Shy-Drager syndrome, and mechanical and traumatic disorders, such as spinal cord disorders, head trauma and stroke. The proteins may also be used as a contraceptive, and for treating coagulation disorders such as haemophilias. The protein and nucleotide sequences with cadherin activity are useful for treating cancer. Other uses for the protein include for inhibiting the growth, infection or function of, or killing, infectious agents such as bacteria, virus, fungi and other parasites, for effecting bodily characteristics such as height, weight, hair colour, effecting biorhythms or cardiac cycles or rhythms, effecting metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors, effecting behavioural characteristics, providing analgesic effects and for treating disorders such as psoriasis.

CC type: 0
XX 537 T: 0 other;[illegible]

RESULT 6
AAZ33929

XX
AC AA733929:

XX 07 DEC 1999 (first entry)

XX amino acid sequence.

DE	Human PRO-2, mRNA
XX	XX: expressed sequence tag; PCR primer; hybridisation;
KW	Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW	probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW	transmembrane protein; ss.

XX
Homo sapiens.XX
10006381-A2

XX
XX

PD TO SEE 1555.
XX
XX

PE 08-MAR-1999;

CC HCSR and for diagnosis of HCSR-related disorders. HCSR and its
CC catalytic or immunogenic fragments are useful for drug screening using
CC libraries of compounds.
XX

	Query Match	12.0%	Score 73;	DB 21;	Length 2099;
	Best Local Similarity	72.9%	Pred. No. 2e-13;		
	Matches 94;	Conservative 0;	Mismatches 35;	Indels 0;	Gaps 0;
QY	85	tgtctactgcacatcatgcataagggcgagtaatctgctcaaacatctttcaacgcacag	144		
Db	1523	tgtctactggcatccagtaaatatagaagcgggggtccgctaaacactctataatgcacag	1582		
QY	145	gcacagagcccccacaaaagagaatattctagcccccataatgtccataaacactgtgttgaga	204		
Db	1583	ggcagttaccccaacagcaaaaataatctgccccaaaatgtcagttgtactgagtttgaga	1642		
QY	205	aaacctacc	213		
Db	1643	aaccacacg	1651		

RESULT 5
AAC59845
AAC59845 standard: DNA: 2116 BP.

XX
XX
AC
AAC59845;
XX
XX

XX
XX

XX secreted protein; human; autoimmune disorder; multiple sclerosis; ulcer;
 DE systemic lupus erythematosus; rheumatoid arthritis; anaemia; stroke;
 KW systemic lupus erythematosus; rheumatoid arthritis; anaemia; stroke;
 KW haematopoiesis regulation; tissue regrowth; wound healing; haemophilia;
 KW Alzheimer's disease; Parkinson's disease; Shy-drager syndrome; cancer;
 KW contractive; infection; growth inhibition; hyperproliferative disorder;
 KW psoriasis; ds.
 KW

XX

[illegible][illegible]

PD
XX
21-SEP-2000.

17-MAR-2000; 2000WO-US0/285.

PR 17-MAR-1999; 99US-0124808.
99US-0124808.
99US-0124808.
99US-0124808.

PR 17-AUG-1999; 99US-0149639.

PR 01-OCT-1999; 99US-0167824.
PR 29-NOV-1999; 99US-0167824.

PR 15-FEB-2000; 2000US-0182/11.

PA (ALPH-) ALPHAGENE INC.

PI Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;

XX
DR WPT: 2000-638211/61.

DR P-PSDB; AAB34744.

PT Novel proteins and polypeptides useful for the treatment of erythematous, autoimmune, infectious, rheumatoid arthritis,

PT cancer, Alzheimer's disease, Parkinson's disease, stroke, anemia and

PT
ulcers
-
XX

PS Claim 124; Page 462; 493pp; English.

This invention relates to 59 human secreted proteins and the nucleotide sequences AAC59788-C59846 and AAB34687-B34745

CC sequences encoding proteins and their encoding nucleotide sequences, and
CC represent the proteins and their encoding nucleotide sequences, and
CC sequences encoding proteins and their encoding nucleotide sequences, and
CC sequences encoding proteins and their encoding nucleotide sequences, and

XX	10-MAR-1998;	98US-0077450.
PR	11-MAR-1998;	98US-0077632.
PR	11-MAR-1998;	98US-0077641.
PR	11-MAR-1998;	98US-0077649.
PR	12-MAR-1998;	98US-0077791.
PR	13-MAR-1998;	98US-0078004.
PR	14-MAR-1998;	98US-0040220.
PR	17-MAR-1998;	98US-0078886.
PR	20-MAR-1998;	98US-0078910.
PR	20-MAR-1998;	98US-0078936.
PR	20-MAR-1998;	98US-0078939.
PR	25-MAR-1998;	98US-0079294.
PR	25-MAR-1998;	98US-0079656.
PR	27-MAR-1998;	98US-0079663.
PR	27-MAR-1998;	98US-0079664.
PR	27-MAR-1998;	98US-0079689.
PR	27-MAR-1998;	98US-0079728.
PR	27-MAR-1998;	98US-0079786.
PR	30-MAR-1998;	98US-0079920.
PR	30-MAR-1998;	98US-0079923.
PR	31-MAR-1998;	98US-0080105.
PR	31-MAR-1998;	98US-0080107.
PR	31-MAR-1998;	98US-0080165.
PR	31-MAR-1998;	98US-0080394.
PR	01-APR-1998;	98US-0080327.
PR	01-APR-1998;	98US-0080328.
PR	01-APR-1998;	98US-0080333.
PR	01-APR-1998;	98US-0080334.
PR	08-APR-1998;	98US-0080409.
PR	08-APR-1998;	98US-0080170.
PR	08-APR-1998;	98US-0080171.
PR	09-APR-1998;	98US-0081195.
PR	09-APR-1998;	98US-0081229.
PR	09-APR-1998;	98US-0081817.
PR	15-APR-1998;	98US-0081838.
PR	15-APR-1998;	98US-0081952.
PR	15-APR-1998;	98US-0081955.
PR	21-APR-1998;	98US-0082568.
PR	21-APR-1998;	98US-0082569.
PR	22-APR-1998;	98US-0082700.
PR	22-APR-1998;	98US-0082804.
PR	23-APR-1998;	98US-0082767.
PR	23-APR-1998;	98US-0082796.
PR	27-APR-1998;	98US-0083336.
PR	28-APR-1998;	98US-0083322.
PR	29-APR-1998;	98US-0083392.
PR	29-APR-1998;	98US-0083495.
PR	29-APR-1998;	98US-0083496.
PR	29-APR-1998;	98US-0083499.
PR	29-APR-1998;	98US-0083500.
PR	29-APR-1998;	98US-0083545.
PR	29-APR-1998;	98US-0083558.
PR	29-APR-1998;	98US-0083558.
PR	29-APR-1998;	98US-0083742.
PR	30-MAY-1998;	98US-0084366.
PR	06-MAY-1998;	98US-0084414.
PR	06-MAY-1998;	98US-0084441.
PR	07-MAY-1998;	98US-0084598.
PR	07-MAY-1998;	98US-0084600.
PR	07-MAY-1998;	98US-0084637.
PR	07-MAY-1998;	98US-0084637.
PR	07-MAY-1998;	98US-0084639.
PR	07-MAY-1998;	98US-0084640.
PR	13-MAY-1998;	98US-0084643.
PR	13-MAY-1998;	98US-0085323.
PR	13-MAY-1998;	98US-0085338.
PR	13-MAY-1998;	98US-0085339.
PR	15-MAY-1998;	98US-0085573.
PR	15-MAY-1998;	98US-0085579.
PR	15-MAY-1998;	98US-0085579.

15-MAY-1998; 98US-0085580.
PR 98US-0085582.
PR 15-MAY-1998; 98US-0085689.
PR 15-MAY-1998; 98US-0085689.
PR 15-MAY-1998; 98US-0085697.
PR 15-MAY-1998; 98US-0085700.
PR 15-MAY-1998; 98US-0085704.
PR 15-MAY-1998; 98US-0086023.
PR 18-MAY-1998; 98US-0086392.
PR 22-MAY-1998; 98US-0086414.
PR 22-MAY-1998; 98US-0086430.
PR 22-MAY-1998; 98US-0086486.
PR 22-MAY-1998; 98US-0087098.
PR 28-MAY-1998; 98US-0087106.
PR 28-MAY-1998; 98US-0087208.
PR 28-MAY-1998; 98US-0094651.
PR 30-JUL-1998; 98US-0094651.
PR 11-SEP-1998; 98US-0100038.
XX
XX (GETH) GENENTECH INC.
PA
XX Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
PI
XX WPI; 1999-551358/46.
XX P-PSDB; AA41690.
DR
XX New secreted and transmembrane polypeptides and their polynucleotides,
XX useful for treating blood coagulation disorders, cancers and cellular
PT adhesion disorders -
PT
PT
PS Claim 2; Fig 19; 530pp; English.
PS
XX The present invention describes secreted and transmembrane polypeptides
XX and their polynucleotides. The nucleotide sequences are useful as
CC sources of probes, primers, for chromosome mapping, and for generation
CC of antisense sequences. They can also be used to create transgenic
CC animals. The proteins can be used to treat a variety of diseases and
CC disorders, depending on their function. Diseases that may be treated
CC include blood coagulation disorders, cancers and cellular adhesion
CC disorders. They may also be used to raise antibodies. AA233891 to
CC AA234338, and AA41685 to AA41774 represent polynucleotide and
CC polypeptide sequence given in the exemplification of the present
CC invention.
CC
XX Sequence 2061 BP; 511 A; 528 C; 491 G; 531 T; 0 other;
SQ

[illegible]

RESULT	7
AAC78471	
ID	AAC78471 standard; cDNA; 2061 BP.
XX	
XX	
AC	AAC78471;
XX	
XX	08-FEB-2001 (first entry)
DT	
XX	
XX	Human PRO329 (UHQ291) nucleotide sequence SEQ ID NO:44.
DE	
XX	
XX	Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;
KW	

us-09-656-668-198.rng

Wed Nov 7 09:21:20 2001

expressed sequence tag; detection; cancer; ss.

XX OS Homo sapiens.

XX PN WO200053756-A2.

XX PD 14-SEP-2000.

XX PF 18-FEB-2000; 2000WO-US04341.

XX PR 08-MAR-1999; 99WO-US05028.

XX PR 12-MAR-1999; 99US-0123957.

XX PR 29-MAR-1999; 99US-0126773.

XX PR 21-APR-1999; 99US-0130232.

XX PR 28-APR-1999; 99US-0131445.

XX PR 14-MAY-1999; 99US-0134287.

XX PR 23-JUN-1999; 99US-0141037.

XX PR 26-JUL-1999; 99US-0145698.

XX PR 29-OCT-1999; 99US-0162506.

XX PR 30-NOV-1999; 99WO-US28313.

XX PR 02-DEC-1999; 99WO-US28551.

XX PR 16-DEC-1999; 99WO-US28565.

XX PR 30-DEC-1999; 99WO-US30095.

XX PR 30-DEC-1999; 99WO-US31243.

XX PR 05-JAN-2000; 2000WO-US00219.

XX PR 06-JAN-2000; 2000WO-US00277.

XX PR 06-JAN-2000; 2000WO-US00376.

XX PA (GETH) GENENTECH INC.

XX PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;

XX PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;

XX PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;

XX PI Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;

XX PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;

XX DR WPI; 2000-611443/58.

XX DR P-PSDB; AAB44246.

XX PT Novel PRO polypeptides and polynucleotides used in detection methods,

XX PT to target bioactive molecules to specific cells, and to modulate

XX PT cellular activities -

XX PS Claim 2; Fig 19; 636pp; English.

XX CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed

XX CC sequence tag) sequences which encode secreted or transmembrane PRO

XX CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic

XX CC activity. The polynucleotides and polypeptides can be used for detecting

XX CC the presence of PRO polypeptides in samples, for linking bioactive

XX CC molecules to cells and for modulating biological activities of cells,

XX CC using the polypeptides for specific targeting. The polypeptide targeting

XX CC can be used to kill the target cells, e.g. for the treatment of cancers.

XX CC The polypeptide pairs provide specific targeting of bioactive molecules

XX CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in

XX CC the isolation of the PRO polynucleotide sequences.

XX CC Sequence 2061 BP; 511 A; 528 C; 491 G; 531 T; 0 other;

XX SQ Query Match 11.7%; Score 71; DB 21; Length 2061;

XX Best Local Similarity 72.4%; Pred. No. 8.7e-13;

XX Matches 92; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 87 ctactggcatcctaatgcagtagggcgagtaatgtgtgtaaacatctttcaacgcacagga 146

DB 1493 ctactggcatcagtaataagaccgaggggtgcgcgtataacatctctaatgcacagg 1552

QY 147 cagagcccccacaaaggaattatctagcccaaatgtccataacactgtgttgagaaa 206

DB 1553 cagtaccaccacagaaaaataatctggcccaaatgtcagttgtactgagtttgagaaa 1612

QY 207 acctacc 213

DB 1613 cccagc 1619

RESULT 8

AAF15763 standard; cDNA; 208 BP.

XX ID AAF15763

XX AC AAF15763;

XX DT 13-MAR-2001 (first entry)

XX DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:198.

XX KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;

XX KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;

XX KW vulnery; gastrointestinal; nephrotropic; antineoplastic; gynaecological;

XX KW antibacterial; gene therapy; neural; immune; reproductive; renal;

XX KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;

XX KW wound; infectious disease; ss.

XX OS Homo sapiens.

XX PN WO200055174-A1.

XX PD 21-SEP-2000.

XX PF 08-MAR-2000; 2000WO-US05988.

XX PR 12-MAR-1999; 99US-0124270.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PA (ROSE/) ROSEN C A.

XX PI Rosen CA, Ruben SM;

XX PI WPI; 2000-587513/55.

XX DR P-PSDB; AAB56560.

XX PT Prostate cancer associated gene sequences, referred to as prostate

XX PT cancer antigens, useful for treatment, prevention, and diagnosis of

XX PT disorders such as prostate cancer -

XX PS Claim 1; Page 736; 2338pp; English.

XX CC AAF15566 to AAF16505 encode the human prostate cancer associated

XX CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.

XX CC The prostate cancer antigens can have neuroprotective, cytostatic,

XX CC cardioactive, immunomodulatory, muscular, vulnery, gastrointestinal,

XX CC nephrotropic, antineoplastic, gynaecological and antibacterial activities,

XX CC and can be used in gene therapy. The prostate cancer antigen

XX CC polynucleotides may be used for detection of prostate cancer, chromosome

XX CC identification, as chromosome markers, and for numerous other diagnostic

XX CC or research purposes. The prostate cancer antigens may be used to treat

XX CC disorders such as neural, immune, muscular, reproductive, proliferative

XX CC gastrointestinal, pulmonary, cardiovascular, renal, and infectious

XX CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to

XX CC AAB57303 represent sequences used in the exemplification of the present

XX CC invention.

XX SQ Sequence 208 BP; 63 A; 38 C; 54 G; 51 T; 2 other;

XX Query Match 11.3%; Score 68.2; DB 21; Length 208;

XX Best Local Similarity 74.4%; Pred. NO. 2.4e-12;

XX Matches 99; Conservative 0; Mismatches 33; Indels 1; Gaps 1;

QY 79 ggaacatgctactggcatcctaatgcagtagaggcagtaatgctgctaaacatctttcaac 138

DB 31 ggggaatgttactggcatcccggtgggtaagagctagggtgctgctagacattctacggt 90

QY 139 gcacaggacagagccccacaaaagagaattattctagcccccaaatgtccataacactgctg 198

ID	AC79962 standard; cDNA; 1583 BP.
XX	
AC	AC79962;
AC	
XX	12-FEB-2001 (first entry)
DT	
DT	
XX	Human secreted protein encoding cDNA for gene 15.
DE	
XX	Secreted protein; human; immunosuppressive; antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective; KW KW antiproliferative; neuroprotective; antibacterial; virucide; fungicide; KW KW neutropic; neuroprotective; antiviral; treatment; autoimmune disease; KW KW ophthalmological; vulnary; gene therapy; cardiovascular disorder; ocular disorder; KW KW hyperproliferative disorder; cardiovascular disorder; infection; skin aging; KW KW cerebrovascular disorder; nervous system disorder; transplantation; ss. KW KW wound healing; epithelial cell proliferation; transplantation; ss.
OS	Homo sapiens.
XX	
PN	W0200058357-A1.
XX	
PD	05-OCT-2000.
PD	
PF	23-MAR-2000; 2000WO-US07723.
XX	
PR	26-MAR-1999; 99US-0126506.
PR	07-JAN-2000; 2000US-0174852.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Rosen CA, Ruben SM, Komatsoulis G;
XX	
DR	WPI: 2000-611704/58.
DR	P-PDSB: AAB45039.
XX	
PT	Nucleic acid molecules encoding human secreted proteins, used in preventing, treating or ameliorating a disorder, e.g. Alzheimer's and Parkinson's diseases and cancers -
PT	
XX	Claim 1a; Page 349; 418pp; English.
PS	
CC	This invention describes novel isolated nucleic acid molecules (I) encoding a human secreted proteins (II) which have immunosuppressive, antiarthritic, antirheumatic, antiproliferative, cytostatic, cardiant, vasotropic, cerebroprotective, neutropic, neuroprotective, antibacterial, virucide, fungicide, ophthalmological and vulnary activity and can be used for gene therapy. (I) and (II) are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. (I) and (II) are also used in diagnosing a pathological condition or susceptibility to a pathological condition. The antibodies to (II) can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioluminoassays or enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. cardiac arrest, of the breast or liver, cardiovascular disorders e.g. cerebral ischemia, angiogenesis, nervous cerebrovascular disorders e.g. Alzheimer's disease, infections caused by bacterial system disorders e.g. viral and fungal diseases, corneal infection. The viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs, before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities.
XX	
SQ	Sequence 1583 BP; 427 A; 370 C; 316 G; 458 T; 12 other;
XX	
XX	Query Match 10.7%; Score 64.6; DB 21; Length 1583;
XX	Best Local Similarity 64.5%; Pred. No. 9.le-11;
XX	Matches 91; Conservative 3; Mismatches 47; Indels 0; Gaps

	Query Match	10.5%	Score 63.4;	DB 21;	Length 2020;
	Best Local Similarity	66.3%;	Pred. No. 2.5e-10;		
	Matches 108;	Conservative 0;	Mismatches 51;	Indels 4;	Gaps 1;
QY	57	tttgattgcacacctgggtgggaacattctacttgctggcatcattgcatagagggcagta	116		
Db	1814	tttgggtgtccactgggtggggg-----tgctattagaatctcttgtagtaaaagccaggga	1869		
QY	117	atgctgtctaaacatctttcaacgcacagcagacagcccaacaagaagaattatctatgcc	176		
Db	1870	atactgtcaaacattctgcaatcacagggatagccccctacaatgaataattatccagac	1929		
QY	177	ccaaatgtccataaacactgctgtgtgagaaaacacctaccgcaggga	219		
Db	1930	ccaatggcgaagtgtagaagggtgtgaaactttgtctctagaa	1972		

Wed Nov 7 09:21:18 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 5, 2001, 23:00:40 ; Search time 126.05 Seconds
(without alignments)
3018.710 Million cell updates/sec

Title: US-09-656-668-198
Perfect score: 606
Sequence: 1 tgagttgccccctaccce.....aagcctgtttgtctgtgac 606

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	606	100.0	606	22	AAF95006 Human ovarian can
2	21	3.5	51	19	AAV36272 Probe used to isol
3	19	3.1	2769	21	AAC76385 Human ORFX ORF1940
4	18	3.0	1427	11	AAQ06902 Sequence encoding
5	18	3.0	1658	21	AAF21367 Human low adenosi
6	18	3.0	1658	21	AAA35245 Human adenosi
7	18	3.0	1674	11	AAQ06901 Sequence encoding
8	18	3.0	2582	18	AAU87428 Clone J422. Homo
9	18	3.0	2635	19	AAV15587 Human B-cell activ
10	18	3.0	2697	20	AAV83816 cDNA encoding a B
11	18	3.0	2775	18	AAU72271 Human B cell surfa

c	12	18	3.0	2775	21	AAA08399 Human RPL05 nucleo
	13	18	3.0	3191	21	AAF21368 Human low adenosi
	14	18	3.0	3191	21	AAAC35246 Human adenosi
	15	18	3.0	1830121	17	AAAT42063 Haemophilus influe
	16	17	2.8	268	21	AAC33704 Arabidopsis thalia
	17	17	2.8	352	20	AAV86186 EST clone J638. H
	18	17	2.8	394	16	AAU26421 Human gene signatu
	19	17	2.8	396	20	AAV87210 EST clone BN99. H
	20	17	2.8	497	21	AAU53869 Arabidopsis thalia
	21	17	2.8	542	21	AAU53133 Human secreted pro
	22	17	2.8	776	18	AAU61702 Rat probasin encod
	23	17	2.8	1329	18	AAU67996 H. pylori cytoplas
	24	17	2.8	1383	18	AAU66164 Mouse interleukin-
	25	17	2.8	1551	21	AAU81117 Human secreted pro
	26	17	2.8	1680	22	AAU23411 Human SEC2 DNA seq
	27	17	2.8	2364	21	AAU55926 cDNA encoding xeno
	28	17	2.8	2601	20	AAU90036 Fervidobacterium p
	29	17	2.8	5921	20	AAU32700 Human glycophospha
	30	17	2.8	8625	21	AAU59470 Human ORFX ORF1389
	31	17	2.8	15512	19	AAU59470 Human ryanodin rec
	32	17	2.8	43795	21	AAU92583 Human DAZ genomic
	33	17	2.8	240825	22	AAU24497 Human PG-3 gene.
	34	17	2.6	166	21	AAU25754 Human secreted pro
	35	16	2.6	264	21	AAU30995 Sequence encoding
	36	16	2.6	269	21	AAU293797 Human secreted pro
	37	16	2.6	317	21	AAU29682 Expressed Sequence
	38	16	2.6	371	14	AAU39834 Human brain expres
	39	16	2.6	371	14	AAU59246 Human secreted pro
	40	16	2.6	375	21	AAU26100 cDNA encoding huma
	41	16	2.6	378	21	AAU78073 Human secreted pro
	42	16	2.6	401	21	AAU03103 Human colon cancer
	43	16	2.6	422	21	AAU80244 Human secreted pro
	44	16	2.6	44	21	AAU01423
	45	16	2.6	484	21	AAU01423

ALIGNMENTS

RESULT 1
AAF95006
ID AAF95006 standard; DNA; 606 BP.
XX
AC AAF95006;
XX
DT 23-MAY-2001 (first entry)
XX
DE Human ovarian cancer associated coding sequence SEQ ID NO: 198.
XX
KW Human, ovarian cancer; vaccine; gene therapy; carcinoma; ds.
XX
OS Homo sapiens.
XX
PN WO200118046-A2.
XX
PD 15-MAR-2001.
XX
PF 08-SEP-2000; 2000WO-US24827.
XX
PR 10-SEP-1999; 99US-0394374.
XX
PR 01-MAY-2000; 2000US-0561778.
XX
PR 15-AUG-2000; 2000US-0640173.
XX
PR 07-SEP-2000; 2000US-0656668.
XX
XX (CORI-) CORIXA CORP.
XX
XX Xu J, Stolk JA;
XX
XX WPI; 2001-211395/21.
XX
XX Isolated polypeptides associated with ovarian carcinomas, and the
XX
XX nucleic acids that encode them, useful for the prevention diagnosis and
XX
XX treatment of ovarian cancers -

XX
XX
PS Claim 18; Page 189; 189pp; English.
XX
CC The present invention provides a number of coding sequences and proteins,
CC the over-expression of which is associated with ovarian carcinoma/cancer.
CC These can be used in the diagnosis, treatment and prevention of ovarian
CC cancer, optionally by gene therapy or in the form of a vaccine. The
CC present sequence is an example of one of these sequences.

SQ XX sequence 606 BP; 175 A; 147 C; 123 G; 161 T; 0 other;
SQ

```

Query Match      100.0%; Score 606; DB 22; Length 606;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 606; Conservative 0; Mismatches 0; Gaps 0;
Indels 0;

```

QY	1	tgagttggcccttaccctccatccagtgaaattttgcatttccataatcctaaagacgctgttctg	60
Db	1	tgagttggcccttaccctccatccagtgaaattttgcatttccataatcctaaagacgctgttctg	60
QY	61	atttcacacctgggtgggaacatgctactggcatactaatgcatagagggcagtaatgc	120
Db	61	atttgcacacctgggtgggaacatgctactggcatactaatgcatagagggcagtaatgc	120
QY	121	tgtcaaacattttcaacgcacagcacagagcccccaaaaagagaattatctagcccca	180
Db	121	tgtcaaacattttcaacgcacagcacagagcccccaaaaagagaattatctagcccca	180
QY	181	atgtccataaacactgctgttgaaaaaacctacgcaggatcttactggcttctcatagta	240
Db	181	atgtccataaacactgctgttgaaaaaacctacgcaggatcttactggcttctcatagta	240
QY	241	agcttgcccttgttgctctctgtagatatataaaataaagacactgccagtcctctcc	300
Db	241	agcttgcccttgttgctctctgtagatatataaaataaagacactgccagtcctctcc	300
QY	301	ctcaacgtcccgagccaggctcaaggcaatttcccaataacagtagaatgaacactaaata	360
Db	301	ctcaacgtcccgagccaggctcaaggcaatttcccaataacagtagaatgaacactaaata	360
QY	361	ttagtttcaaaactcagcaactagagaatgaaccaacctctcgttggcctgggacgtg	420
Db	361	ttagtttcaaaactcagcaactagagaatgaaccaacctctcgttggcctgggacgtg	420
QY	421	tctagtttttagcattgaaagttttcaggtttcaggaaagccctcaggcctgggctgctgg	480
Db	421	tctagtttttagcattgaaagttttcaggtttcaggaaagccctcaggcctgggctgctgg	480
QY	481	tcaccctagcagctgagggaactcttaatacagaattagctctttgtgcactggagatgaa	540
Db	481	tcaccctagcagctgagggaactcttaatacagaattagctctttgtgcactggagatgaa	540
QY	541	tatactttaattgtgaacgtgaaaacatctataaacatctactgaagcctgttcttgt	600
Db	541	tatactttaattgtgaacgtgaaaacatctataaacatctactgaagcctgttcttgt	600

Db	541	tatactttaatgtgaacatgtgaaacatctctataaacatctctctcgaagccgctctctcgtg
Qy	601	ctgcac 606
Db	601	ctgcac 606
RESULT 2		
AAV36272/c		
ID	AAV36272 standard; DNA; 51 BP.	
XX		
AC	AAV36272;	
XX		
DT	08-SEP-1998 (first entry)	
XX		
DE	Probe used to isolate DNA encoding Elongin A homologue of <i>C. elegans</i> .	
XX		
DE	Elongin A; rat; elongation activity; stimulation; RNA polymerase II;	
XX	Elongin B; Elongin C; regulation; transcription activity;	
KW		
KW		
KW		

XX	RNA polymerase; in vitro transcriptional assay, probe, etc.
XX	Synthetic.
OS	Caenorhabditis elegans.
OS	
PN	W09814574-A1.
XX	
XX	09-APR-1998.
PD	
XX	02-OCT-1997; 97NO-US17992.
XX	
XX	04-OCT-1996; 96US-0725459.
PR	
XX	{OKLA-} OKLAHOMA MEDICAL RES FOUND.
PA	
XX	Conaway JW, Conaway RC;
PI	
XX	WPI; 1998-286420/25.
DR	
XX	Isolated Elongin A and C and fragments - used for regulating
XX	transcriptional activation of RNA polymerase, particularly for use
PT	in in vitro assays and systems
PT	
XX	Example 7; Page 44; 185pp; English.
PS	
XX	The present sequence represents a probe used to isolate DNA encoding an
CC	Elongin A homologue of Caenorhabditis elegans. Elongin A stimulates
CC	the elongation activity of RNA polymerase II. Fragments of the rat
CC	Elongin A (AAW60772-76) have been found to each possess 50% or more of
CC	the elongation activation activity exhibited by the wild type protein.
CC	The Elongin A sequence critical for binding to Elongins B and C is also
CC	found within these fragments. The Elongin polypeptides and fragments can
CC	be used for regulating the transcriptional activity of RNA polymerase.
CC	They can be used in in vitro transcriptional assays or systems.
XX	Sequence 51 BP; 13 A; 9 C; 17 G; 12 T; 0 other;
SO	

```
Query Match          3.5%; Score 21; DB 19; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 579 atctactgaagcctgttcttg 599
 |
Db 44 ATCTACTGAAGCCTGTTCTTG 24

RESULT	3
AAC76385/c	
ID	AAC76385 standard; cDNA; 2769 BP.
XX	
XX	AAC76385;
AC	
XX	
XX	08-FEB-2001 (first entry)
DT	
XX	
XX	

DE	Human ORFX ORF1940 polynucleotide sequence
XX	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW	vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW	immunostimulant; thrombolytic; coagulant; vasotonic; antidiabetic;
KW	hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW	antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW	antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW	cholesterol ester storage; systemic lupus erythematosus; infection;
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW	thrombosis; contraceptive, ss.
OS	Homo sapiens.

us-09-656-668-198.oli.rng

Wed Nov 7 09:21:18 2001

```

XX FH Key Location/Qualifiers
XX CDS 33..818
XX FT /*tag= a
XX FT 33..92
XX FT /*tag= b
XX FT 93..818
XX FT /*tag= c
XX FT
XX FT
XX PN EP403114-A.
XX PD 19-DEC-1990.
XX PF 31-MAY-1990; 90EP-0305928.
XX PR 14-MAR-1990; 90US-0493588.
XX PR 15-JUN-1989; 89US-0366910.
XX PR 13-OCT-1989; 89US-0421201.
XX PA (IMMU-) IMMUNEX CORP.
XX PI Park LS, Goodwin RG;
XX DR WPI; 1990-377843/51.
XX DR P-PSDB; AAR08330.
XX XX Mammalian interleukin-7 receptor DNA, protein and analogues -
XX PT used in therapy, diagnosis, assay and antibody production
XX PS Claim 4; Fig 3; 28pp; English.
XX XX IL-7R gene product may be used in immunoassay or for affinity
XX CC purification eg. IL-7R, IL-1 or IL-1R. May also be useful in
XX CC regulation of immune response specifically of IL-7 in mammals
XX CC esp. humans.
XX XX Sequence 1427 BP; 438 A; 313 C; 313 G; 362 T; 1 other;
XX SQ

Query Match 3.0%; Score 18; DB 11; Length 1427;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 374 ctacgaactagaagaat 391
|||||
Db 924 ctacgaactagaagaat 941

RESULT 5
AAF21367
ID AAF21367 standard; DNA; 1658 BP.
XX AC
XX AAF21367;
XX DT 14-MAR-2001 (first entry)
XX DE Human low adenosine antisense oligonucleotide related sequence #2934.
XX DE Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX KW
XX OS Homo sapiens.
XX XX
XX PN WO200062736-A2.
XX XX

```

```

XX PN WO200058473-A2.
XX PD 05-OCT-2000.
XX PF 31-MAR-2000; 2000WO-US08621.
XX PR 31-MAR-1999; 99US-0127607.
XX PR 02-APR-1999; 99US-0127636.
XX PR 05-APR-1999; 99US-0127728.
XX PR 30-MAR-2000; 2000US-0540763.
XX PA (CURA-) CURAGEN CORP.
XX PI Shimkets RA, Leach M;
XX DR WPI; 2000-602362/57.
XX DR P-PSDB; AAB42176.
XX XX Novel nucleic acids and peptides derived from open reading frame X,
XX PT useful for treating e.g. cancers, proliferative disorders,
XX PT neurodegenerative disorders and cardiovascular disease -
XX PS Claim 5; Page 3025-3027; 5507pp; English.
XX XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
XX CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
XX CC osteostimulant; cardiant; thrombotic; coagulant; vasotropic;
XX CC immunosuppressive; dermatological; immunosuppressive;
XX CC antidiabetic; hypotensive; antihypertensive; antihypertensive;
XX CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX CC antithyroid; and antianemic. The sequences can be used for determining
XX CC the presence of or predisposition to, or preventing or treating
XX CC pathological conditions associated with an ORFX-associated disorder. The
XX CC nucleic acids can be used to express ORFX proteins in gene therapy
XX CC vectors. The proteins and nucleic acids may be used to treat cancers,
XX CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX CC graft vs host disease, cardiovascular disease, diabetes mellitus,
XX CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
XX CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX XX Sequence 2769 BP; 604 A; 789 C; 772 G; 603 T; 1 other;
XX SQ

Query Match 3.1%; Score 19; DB 21; Length 2769;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 425 agtttagcattgaagtt 443
|||||
Db 1419 AGTTTAGCATTGAAGTT 1401

RESULT 4
AAQ06902
ID AAQ06902 standard; DNA; 1427 BP.
XX AC
XX AAQ06902;
XX DT 04-MAR-1991 (first entry)
XX DE Sequence encoding human IL-7 receptor clone H6.
XX XX
XX XX Interleukin-7 receptor; immune response; pre-B cell growth factor;
KW Lymphopoietin; ds.
XX PN
XX OS Homo sapiens.

```

Wed Nov 7 09:21:18 2001

allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
respiratory distress syndrome; pain; cystic fibrosis; emphysema;
pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

KW

Homo sapiens.
WO200009525-A2.

OS
XX
PN

24-FEB-2000.
03-AUG-1999; 99WO-US17712.
03-AUG-1998; 98US-0095212.

XX
XX
XX

(UYEC-) UNIV EAST CAROLINA.
Nyce JW;

XX
XX
XX

WPI; 2000-205971/18.
New antisense oligonucleotides useful for treating e.g. pulmonary
vasoconstriction, inflammation, allergies, asthma, hypertension,
bronchitis, emphysema, respiratory distress syndrome, ischemia or
cancers -

XX
XX
XX

Disclosure; Page 1284-1285; 1343pp; English.
The present invention describes a new composition comprising an
antisense oligonucleotide (ON) with low adenosine (up to 15%), which
targets nucleic acids involved in bronchoconstriction, allergies, and/or
inflammation. The ON can have antiinflammatory, antiallergic,
antiasthmatic, cytostatic and analgesic activities. The compositions are
useful for the treatment of diseases associated with inflammation,
impaired airways, including lung disease and diseases whose secondary
effects afflict the lungs of a subject. They can be used for treating
e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, cystic
asthma, impaired respiration, respiratory distress syndrome, pain, chronic
pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
carcinomas, and cancers which may metastasize to the lungs, including
breast and prostate cancer. The reduction of the adenosine content of
the ONs reduces side effects. The A-containing ONs break down with the
release of deoxyadenosine which activates adenosine receptors causing the
bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
nucleotide sequences given in the sequence listing from the present
invention, which correspond to SEQ ID NO:1 to 185, and then the last
185 sequences are also called SEQ ID NO:1 to 185, but the sequences
differ from the previously named sequences. SEQ ID NO:11 to 1680
(AAA32323 to AAA3992) are specifically claimed ONs from the present
invention. N.B. Sequences given in the disclosure of the present
invention do not match up with their corresponding SEQ ID NO: sequences
given in the sequence listing.

XX
XX
XX

Query Match 3.0%; Score 18; DB 21; Length 1658;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX
XX
XX

374 ctcagcaactagaagaat 391
|||||

QY

1008 ctcagcaactagaagaat 1025

DB

RESULT 7
AAQ06901
ID AAQ06901 standard; CDNA; 1674 BP.

XX
XX
XX

AAQ06901

XX
XX
XX

26-OCT-2000.
24-MAR-2000; 2000WO-US08020.
06-APR-1999; 99US-0127958.

PD
XX
PF

(UYEC-) UNIV EAST CAROLINA.
(NYCE/) NYCE J W.

XX
XX
XX

WPI; 2000-679539/66.
Low adenosine (A) content antisense oligonucleotides which do not
trigger adenosine receptors during metabolism, useful e.g. for treating
cancers and respiratory obstructions -

XX
XX
XX

Disclosure; Page 1370; 1592pp; English.
The present invention describes low adenosine (A) content antisense
oligonucleotides and compositions (I) comprising them. In the antisense
oligonucleotides the A is replaced by a 'universal' or alternative base.
(I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
The antisense oligonucleotides and (I) can be used to down-regulate the
expression and/or activity of target polypeptides associated with
lung/respiratory disorders and malignancies, such as stimulating and
activating peptide factors and transmitters, transcription factors,
immunoglobulins and antibodies, antibody receptors, cytokines and
chemokines, endogenously produced specific and non-specific enzymes.
binding proteins, adhesion molecules and their receptors, cytokine and
chemokine receptors, adenosine receptors, bradykinin receptors, central
nervous system (CNS) and peripheral nervous and non-nervous system peptide
receptors, CNS and peripheral nervous and non-nervous system peptide
transmitters, defensins, growth factors, vasoactive peptides and
receptors, binding proteins and malignancy associated proteins. The
antisense oligonucleotides may be used in this way to treat disorders
including respiratory obstruction (especially pulmonary obstruction)
and/or bronchoconstriction and/or lung inflammation, allergy(ies)
and/or surfactant hypoproduction which are associated with a disease or
condition selected from pulmonary vasoconstriction, inflammation,
allergies, asthma, impaired respiration, respiratory distress syndrome
(RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
pulmonary transplantation rejection, pulmonary infections, bronchitis,
and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
fragments and antisense oligonucleotides used in the exemplification of
the present invention.

XX
XX
XX

Query Match 3.0%; Score 18; DB 21; Length 1658;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX
XX
XX

374 ctcagcaactagaagaat 391
|||||

QY

1008 ctcagcaactagaagaat 1025

DB

RESULT 6
AAA35245
ID AAA35245 standard; DNA; 1658 BP.

XX
XX
XX

AAA35245

XX
XX
XX

28-JUL-2000 (first entry)
Human adenosine receptor related polynucleotide 2nd SEQ ID NO:119.

XX
XX
XX

Human; adenosine receptor; low adenosine antisense oligonucleotide;
phosphorothioate; impaired respiration; inflammation; allergy;

XX
XX
XX

us-09-656-668-198.oli.rng

Wed Nov 7 09:21:18 2001

```

XX PS Claim 2; Page 10-13; 15pp; Japanese.
XX The present sequence encodes a B cell surface protein. The protein
CC promotes growth of B cells as well as inhibition and induction of B
CC cell apoptosis. The protein and the DNA are useful in the treatment
CC and as a diagnostic agent for B-cell antibody production related
CC diseases e.g. autoimmune diseases, allergic diseases, asthma and
CC atopic dermatitis.
XX SQ Sequence 2697 BP; 757 A; 680 C; 533 G; 727 T; 0 other;

Query Match 3.0%; Score 18; DB 20; Length 2697;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 ctgccagtcctccctc 303
DB 2419 CTGCCAGTCCTCCCTC 2402

RESULT 11
AAV83816/C
ID AAV83816 standard; cDNA; 2697 BP.
XX AC AAV83816;
XX DT 03-MAR-1999 (first entry)
XX DE cDNA encoding a B cell surface protein.
XX KW B cell surface protein; growth; inhibition; induction; B cell apoptosis;
XX KW autoimmune disease; allergic disease; asthma; atopic dermatitis; ds.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 143..2128
XX FT /*tag= a
XX FT sig_peptide 143..203
XX FT /*tag= b
XX FT mat_peptide 204..2125
XX FT /*tag= c
XX JP10313870-A.
XX PN
XX PD 02-DEC-1998.
XX PF 22-MAY-1997; 97JP-0132592.
XX PR 22-MAY-1997; 97JP-0132592.
XX PA (MITU ) MITSUBISHI CHEM CORP.
XX DR WPI; 1999-084647/08.
XX DR P-PSDB; AAW87556.
XX PT New DNA sequence encoding B cell surface protein - useful as
XX PT diagnostic agent and for treating auto-immune diseases, allergic
XX PT diseases, asthma and atopic dermatitis

Query Match 3.0%; Score 18; DB 19; Length 2635;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 ctgccagtcctccctc 303
DB 2373 CTGCCAGTCCTCCCTC 2356

RESULT 10
AAV83816/C
ID AAV83816 standard; cDNA; 2697 BP.
XX AC AAV83816;
XX DT 03-MAR-1999 (first entry)
XX DE cDNA encoding a B cell surface protein.
XX KW B cell surface protein; growth; inhibition; induction; B cell apoptosis;
XX KW autoimmune disease; allergic disease; asthma; atopic dermatitis; ds.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 143..2128
XX FT /*tag= a
XX FT sig_peptide 143..203
XX FT /*tag= b
XX FT mat_peptide 204..2125
XX FT /*tag= c
XX JP10313870-A.
XX PN
XX PD 02-DEC-1998.
XX PF 22-MAY-1997; 97JP-0132592.
XX PR 22-MAY-1997; 97JP-0132592.
XX PA (MITU ) MITSUBISHI CHEM CORP.
XX DR WPI; 1999-084647/08.
XX DR P-PSDB; AAW87556.
XX PT New DNA sequence encoding B cell surface protein - useful as
XX PT diagnostic agent and for treating auto-immune diseases, allergic
XX PT diseases, asthma and atopic dermatitis

Query Match 3.0%; Score 18; DB 20; Length 2697;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 ctgccagtcctccctc 303
DB 2419 CTGCCAGTCCTCCCTC 2402

RESULT 11
AAV72271/C
ID AAV72271 standard; cDNA; 2775 BP.
XX AC AAV72271;
XX DT 30-SEP-1997 (first entry)
XX DE Human B cell surface antigen cDNA.
XX KW B cell surface antigen; gene vaccine; apoptosis; dexamethasone;
XX KW irradiation; ss.
XX OS Homo sapiens.
XX PN WO9714813-A1.
XX PD 24-APR-1997.
XX PF 16-OCT-1996; 96WO-US16527.
XX PR 19-OCT-1995; 95US-0005642.
XX PA (SMIK ) SMITHKLINE BEECHAM CORP.
XX PI Marshall L, Roshak A;
XX DR WPI; 1997-245128/22.
XX CDNA clone encoding human B cell surface antigen - used in gene
XX vaccine and to prevent apoptosis resulting from dexamethasone
XX treatment and irradiation
XX Claim 1; Page 16-17; 22pp; English.
XX A CDNA clone (AAV72271) codes for a novel human B cell surface
XX antigen that is involved in B cell proliferation and protection from
XX dexamethasone- and irradiation-induced apoptosis. A plasmid
XX containing the CDNA clone was deposited as ATCC 69902. To isolate
XX the clone, CDNA libraries were constructed in the Stratagene
XX UniZapXR lambda phage vector and immediately amplified. Plasmid
XX pBluescriptSK+ (contained within UniZap phage) was excised by co-
XX infection of XL-1 Blue MRF' cells with helper phage and an aliquot
XX of the amplified library. The resulting phagemid stock was used to
XX infect SOLR cells which were then plated on LB+ ampicillin plates.
XX DNA from individual colonies was sequenced. The insert is in the
XX EcoRI(5')/XhoI(3') sites of Bluescript SK+. The isolated clone can
XX be used to produce recombinant antigen, in methods for detecting of
XX mutated forms of the human gene antigen, and in the development of
XX genetic vaccines.

```

Wed Nov 7 09:21:18 2001

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

SQ Sequence 2775 BP; 792 A; 694 C; 565 G; 724 T; 0 other;

Query Match 3.0%; Score 18; DB 18; Length 2775;

Best Local Similarity 100.0%; Pred. No. 13; 0; Indels 0; Gaps 0;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 ctgccagtcctccctc 303

DB 2417 CTGCCAGTCCCTCCCTC 2400

RESULT 12

AAA08399/C

ID AAA08399 standard; DNA; 2775 BP.

XX

AC AAA08399;

XX

DT 12-JUL-2000 (first entry)

XX

DE Human RP105 nucleotide sequence SEQ ID NO:2.

XX

KW Human; RP105; agonist; antagonist; allergy; allergic disease; identification; antiasthmatic; dermatological; antiinflammatory; cytostatic; immunosuppressive; neuroprotective; antianaemic; antidiabetic; asthma; atopic dermatitis; B cell neoplasm; chronic lymphocyte leukaemia; hairy cell leukaemia; myeloma; polymorphic leukaemia; autoimmune disease; rheumatoid arthritis; systemic lupus erythematosus; multiple sclerosis; diabetes; acquired haemolytic anaemia; ds.

XX

OS Homo sapiens.

XX

PN WO200012130-A1.

XX

PD 09-MAR-2000.

XX

XX 24-AUG-1999; 99WO-US19336.

XX

XX 27-AUG-1998; 98US-0098030.

XX

PA (SMIK) SMITHKLINE BEECHAM CORP.

XX

PA (SMIK) SMITHKLINE BEECHAM PLC.

XX

XX Harrop JA, Holmes SD, Roshak AK;

XX

PI WPI; 2000-375617/32.

XX

DR P-PSDB; AAY82527.

XX

XX New anti-murine B cell antigen RP105 antibody for treating or preventing allergic disease such as asthma, B cell neoplasms, for example leukemias, and autoimmune diseases such as systemic lupus erythematosus.

XX

PS Example; Page 26-28; 30pp; English.

XX

XX The present invention describes an anti-murine B cell antigen RP105 antibody (I). The present sequence encodes human RP105, which is used in the exemplification of the present invention. (I) has antiasthmatic, dermatological, antiinflammatory, cytostatic, immunosuppressive, neuroprotective, antianaemic and antidiabetic activities. (I) can be used for in the treatment and prevention of allergic disease, including asthma and atopic dermatitis, B cell neoplasms, including chronic lymphocyte leukaemia, hairy cell leukaemia, polymorphic leukaemia, myelomas, autoimmune diseases such as systemic lupus erythematosus, rheumatoid arthritis, multiple sclerosis, acquired haemolytic anaemia and diabetes.

XX

SQ Sequence 2775 BP; 792 A; 694 C; 565 G; 724 T; 0 other;

Query Match 3.0%; Score 18; DB 21; Length 2775;

Best Local Similarity 100.0%; Pred. No. 13;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 ctgccagtcctccctc 303

DB 2417 CTGCCAGTCCCTCCCTC 2400

RESULT 13

AAF21368

ID AAF21368 standard; DNA; 3191 BP.

XX

AC AAF21368;

XX

DT 14-MAR-2001 (first entry)

XX

DE Human low adenine antisense oligonucleotide related sequence #2935.

XX

KW Low adenine antisense oligonucleotide; phosphorothioate; allergy; human; airway disorder; bronchoconstriction; lung inflammation; surfactant depletion; respiratory; bronchodilator; antiinflammatory; immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic; respiratory obstruction; pulmonary obstruction; impeded respiration; surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS; respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis; pulmonary hypertension; emphysema; pulmonary transplantation rejection; chronic obstructive pulmonary disease; pulmonary infection; bronchitis; cancer; ss.

XX

OS Homo sapiens.

XX

PN WO200062736-A2.

XX

PD 26-OCT-2000.

XX

PF 24-MAR-2000; 2000WO-US08020.

XX

PR 06-APR-1999; 99US-0127958.

XX

PA (UYEC-) UNIV EAST CAROLINA.

XX

PA (NYCE/) NYCE J W.

XX

PI Nyce JW;

XX

WPI; 2000-679539/66.

XX

XX Low adenine (A) content antisense oligonucleotides which do not trigger adenine receptors during metabolism, useful e.g. for treating cancers and respiratory obstructions.

XX

PS Disclosure; Page 1370-1371; 1592pp; English.

XX

XX The present invention describes low adenine (A) content antisense oligonucleotides and compositions (I) comprising them. In the antisense oligonucleotides the A is replaced by a 'Universal' or alternative base. (I) can have respiratory, bronchodilator, antiinflammatory, analgesic, immunosuppressive, antiasthmatic, hypotensive and cytostatic activities. The antisense oligonucleotides and (I) can be used to down-regulate the expression and or activity of target polypeptides associated with lung/respiratory disorders and malignancies, such as stimulating and activating peptide factors and transmitters, transcription factors, immunoglobulins and antibodies, antibody receptors, cytokines and chemokines, endogenously produced specific and non-specific enzymes, binding proteins, adhesion molecules and their receptors, cytokine and chemokine receptors, adenine receptors, bradykinin receptors, central nervous system (CNS) and peripheral nervous and non-nervous system peptide receptors, CNS and peripheral nervous and non-nervous system peptide transmitters, defensins, growth factors, vasoactive peptides and receptors, binding proteins and malignancy associated proteins. The antisense oligonucleotides may be used in this way to treat disorders including respiratory obstruction (especially pulmonary obstruction and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or surfactant hypoproduction which are associated with a disease or condition selected from pulmonary vasoconstriction, inflammation,

Wed Nov 7 09:21:18 2001

CC allergies, asthma, impeded respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.

XX Sequence 3191 BP; 947 A; 674 C; 663 G; 907 T; 0 other;

Query Match 3.0%; Score 18; DB 21; Length 3191;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 374 ctcagcaactagaagaat 391
 |||||
 Db 2541 ctcagcaactagaagaat 2558

RESULT 14
 AAA35246
 ID AAA35246 standard; DNA; 3191 BP.

XX AC AAA35246;

XX DT - 28-JUL-2000 (first entry)

XX Human adenosine receptor related polynucleotide 2nd SEQ ID NO:120.
 DE Human; adenosine receptor; low adenosine antisense oligonucleotide;
 XX phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antiasthmatic; cyostatic; analgesic; impaired airway;
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

XX Homo sapiens.
 XX WO200009525-A2.

XX 24-FEB-2000.

XX 03-AUG-1999; 99WO-US17712.

XX 03-AUG-1998; 98US-0095212.

XX (UYEC-) UNIV EAST CAROLINA.

XX Nyce JW;

XX WPI; 2000-205971/18.

XX New antisense oligonucleotides useful for treating e.g. pulmonary
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
 PT cancers.

XX Disclosure; Page 1285-1286; 1343pp; English.

XX The present invention describes a new composition comprising an
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
 CC inflammation. The ON can have antiinflammatory, antiallergic,
 CC antiasthmatic, cyostatic and analgesic activities. The compositions are
 CC useful for the treatment of diseases associated with inflammation,
 CC impaired airways, including lung disease and diseases whose secondary
 CC effects afflict the lungs of a subject. They can be used for treating
 CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
 CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive

CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
 CC carcinomas, and cancers which may metastasise to the lungs, including
 CC breast and prostate cancer. The reduction of the adenosine content of
 CC the ONs reduces side effects. The A-containing adenosine receptors causing the
 CC release of deoxyadenosine which activates adenosine receptors causing the
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680
 CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
 CC invention. N.B. Sequences given in the disclosure of the present
 CC invention do not match up with their corresponding SEQ ID NO: sequences
 CC given in the sequence listing.

XX Sequence 3191 BP; 947 A; 674 C; 663 G; 907 T; 0 other;

Query Match 3.0%; Score 18; DB 21; Length 3191;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 374 ctcagcaactagaagaat 391
 |||||
 Db 2541 ctcagcaactagaagaat 2558

RESULT 15

AAAT42063
 ID AAT42063 standard; DNA; 1830121 BP.

XX AC AAT42063;

XX 14-SEP-1999 (first entry)

DE Haemophilus influenzae complete genome sequence.

XX Genome; bacterium; Haemophilus influenzae; computer readable medium;
 KW expression modulating fragment; regulation; gene expression; vector;
 KW organism; open reading frame; ORF; ds.

XX Haemophilus influenzae.

XX WO9633276-A1.

XX 24-OCT-1996.

XX 22-APR-1996; 96WO-0505320.

XX 07-JUN-1995; 95US-0487429.

XX 21-APR-1995; 95US-0426787.

XX 07-JUN-1995; 95US-0476102.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (UYJO) UNIV JOHNS HOPKINS.

XX Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;

XX WPI; 1996-485782/48.

XX Haemophilus influenzae Rd genome recorded on computer readable
 PT medium - useful for identifying commercially important nucleic acid
 PT fragments by homology searching

XX Claim 1; Page 77.2-77.1091; 1291pp; English.

XX This sequence represents the complete genome sequence of the bacterium
 CC Haemophilus influenzae strain Rd. The invention relates to a computer
 CC readable medium (CRM) having recorded upon it the complete H. influenzae
 CC nucleotide sequence (1), a representative fragment of (1) or a nucleotide
 CC sequence at least 99% identical to (1). By providing the full-length
 CC genomic sequence in a computer readable form, it is possible to identify
 CC commercially important nucleic acid fragments and expression modulating

Wed Nov 7 09:21:18 2001

CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
 CC regulate the expression of a nucleic acid molecule. Vectors and altered
 CC organisms comprising the predicted ORFs can be used to produce any of the
 CC polypeptide fragments of the H. influenzae Rd genome.

XX
 SQ Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;

Query Match 3.0%; Score 18; DB 17; Length 1830121;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 580 tctactgaagcctgttct 597
 |||||
 Db 408370 tctactgaagcctgttct 408387

Search completed: November 5, 2001, 23:48:56
 Job time: 2896 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

ow nucleic - nucleic search. using sw model

Run on: November 5, 2001, 22:16:40 ; Search time 87.77 Seconds
(without alignments)
1563.697 Million cell updates/sec

Title: US-09-656-668-198
Perfect score: 606
sequence: 1 tgaattggcccttaccc.....agcctgttctgtctgcac 606

Scoring table: IDENTITY_NUC
Gapox 10.0 . Gapext 1.0

Searched: 351203 seqs, 113238999 residues
Number of bits satisfying chosen parameters: 702406

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 s

```
Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCRTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:
```

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			ID	Description	
	Score	Match	Length			
1	82.4	13.6	80246	4	US-09-078-294-4	Sequence 4, Appli
2	82.4	13.6	80595	4	US-09-078-294-3	Sequence 3, Appli
c 3	64.8	10.7	7505	4	US-09-078-294-13	Sequence 13, Appli
c 4	60.4	10.0	6314	1	US-08-211-430-1	Sequence 1, Appli
5	46.8	7.7	279	3	US-08-211-430-1	Sequence 120, App
6	41	6.8	1750	4	US-09-157-177-120	Sequence 34, Appli
c 7	38.6	6.4	11288	4	US-09-276-531-34	Sequence 1, Appli
c 8	38.6	6.4	11288	4	US-08-646-301A-1	Sequence 4, Appli
c 9	36.6	6.0	3176	2	US-08-481-968A-4	Sequence 17, Appli
c 10	36.6	6.0	3176	2	US-08-910-733-17	Sequence 17, Appli
11	35.6	5.9	11531-1	1	US-08-068-945A-1	Sequence 1, Appli
12	35.6	5.9	11531	1	US-08-442-806-1	Sequence 3, Appli
13	35.4	5.8	1481	3	US-09-423-323-3	Sequence 31, Appli
14	35.4	5.8	3073	2	US-07-688-352C-31	Sequence 31, Appli
15	35.4	5.8	3073	2	US-08-474-379C-31	Sequence 31, Appli
16	35.4	5.8	3073	3	US-09-146-249A-31	Sequence 31, Appli
17	35.4	5.8	3073	3	US-08-206-188B-31	Sequence 31, Appli
18	35.4	5.8	3073	5	PCR-US91-02714-30	Sequence 30, Appli
19	34.4	5.7	246240	2	US-08-724-394A-20	Sequence 20, Appli
20	34.4	5.7	246240	2	US-08-724-394A-21	Sequence 21, Appli
21	34.4	5.7	246240	2	US-08-724-394A-22	Sequence 22, Appli
c 22	34	5.6	2218	2	US-08-421-044-1	Sequence 1, Appli
c 23	32.8	5.4	8353	3	US-08-611-587-1	Sequence 1, Appli
24	31.8	5.2	1524	4	US-08-840-767-3	Sequence 3, Appli
c 25	31.2	5.1	5427	3	US-09-009-913-2	Sequence 2, Appli
c 26	31.2	5.1	5510	3	US-09-009-913-3	Sequence 3, Appli
			5253	3	US-09-008-913-4	Sequence 4, Appli

Sequence 17, Appl
Sequence 12, Appl
Sequence 4, Appl
Sequence 2, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 5, Appl
Sequence 5, Appl
Sequence 5, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 1, Appl
Sequence 97, Appl
Sequence 3, Appl
Sequence 5, Appl
Sequence 2, Appl

ALIGNMENTS

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

[illegible]

RESULT 2
US-09-078-294-3
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.

Wed Nov 7 09:21:20 2001

```

; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 80595
; TYPE: DNA
; ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3

Query Match      13.6%; Score 82.4; DB 4; Length 80595;
Best Local Similarity 63.8%; Pred. No. 4.2e-17;
Matches 125; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 72 tgggtgggaacatgctactggtcattatgcatagagggcagtaatgctgtaaacatc 131
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66788 tgagtgggtgggtgctactggtcattatgcatagagggcagtaatgctgtaaacatc 66847

QY 132 tttaacacagcagcagagcccccacacaaagagaattatctagcccaaatgtccataac 191
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66848 ccgcaagcacagggcagtcgccgcacacaaagagaattatctagcccaaatgtccataac 66907

QY 192 actgctgttgagaaacctaccgcaggatcttacttggtggttcataaggtaagcttgccttt 251
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db .66908 gccaaagtggaaacctcattctagcttcccttctctacgttcttaatacaactgtt 66967

QY 252 gtcttggtctctgttag 267
    ||||| |||||
Db 66968 gtctcttcagcattag 66983

RESULT 3
US-09-078-294-13/c
; Sequence 13, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 7505
; TYPE: DNA
; ORGANISM: BAC-F2 contig 9
US-09-078-294-13

Query Match      10.7%; Score 64.8; DB 4; Length 7505;
Best Local Similarity 72.1%; Pred. No. 1e-11;
Matches 98; Conservative 0; Mismatches 37; Indels 1; Gaps 1;

QY 76 tggggaacatgctactggtcattatgcatagagggcagtaatgctgtaaacatcttc 135
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4460 TTGGGAATTTGCTACAGACATCTAATAATTAGAGGCTAGGGAATTTGTTAAATGTCCTAC 4401

QY 136 aacgcacagcagagagcccccacacaaagagaattatctagcccaaatgtccataacactg 195
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4400 AGTGCACATGACAGCCACCATAAACAGAGAATTGTATAGCCCAAAATGT-CATTAGTGTG 4342

QY 196 ctgttgagaacacctta 211
    ||||| ||||| |||||
Db 4341 ACGTTGAGAAAAATATA 4326

RESULT 4
US-08-211-430-1/c
; Sequence 1, Application US/08211430
; Patent No. 5763166
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: NUCLEIC SEQUENCE OF THE GENE ASSOCIATED WITH
; TITLE OF INVENTION: X CHROMOSOME LINKED KALLMANN SYNDROME, CORRESPONDING
; TITLE OF INVENTION: PEPTIDE SEQUENCES, DIAGNOSTIC APPLICATIONS.
; NUMBER OF SEQUENCES: 32
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/211,430
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6314 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL LINE: foetal brain cells
US-08-211-430-1

Query Match      10.0%; Score 60.4; DB 1; Length 6314;
Best Local Similarity 62.9%; Pred. No. 3e-10;
Matches 127; Conservative 0; Mismatches 71; Indels 4; Gaps 2;

QY 15 taccctccatccagtgaaatttgcaattctcctaagacggtgttttgattgtcacacctgg 74
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5428 TACCATCCCTCAAGGGGTAGCTGGCANAATATCTGGAGACATTTTGGTTGTCTCAACTTG- 5368

QY 75 gtggggaacatgctactggtcattatgcatagagggcagtaatgctgtaaacatcttt 134
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5369 --GGGAGGGTGCACCTGTCATCTGGTGGAGGAGAAAGTGTGGATGCTACTTAA-ACCTA 5313

QY 135 caacycacagcagagcccccacacaaagagaattatctagcccaaatgtccataacact 194
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5312 CAGTGCAACAAGAACACACACACAGAGAACTATCCAGCCCCCAAAAGTCAATAGTGT 5253

QY 195 gctgttgagaaacctaccgca 216
    ||||| ||||| ||||| |||||
Db 5252 GAGGTTGAGATGAACCTCTGTA 5231

RESULT 5
US-09-157-177-120
; Sequence 120, Application US/09157177
; Patent No. 6090558
; GENERAL INFORMATION:
; APPLICANT: Butler, John M.
; APPLICANT: Li, Jia
; APPLICANT: Monforte, Joseph A.
; APPLICANT: Becker, Christopher H.
; TITLE OF INVENTION: DNA TYPING BY MASS SPECTROMETRY WITH POLYMORPHIC DNA
; TITLE OF INVENTION: REPEAT MARKERS
; FILE REFERENCE: GTR-017/GTR017P
; CURRENT APPLICATION NUMBER: US/09/157,177
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: 60/059,415
; EARLIER FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 120
; LENGTH: 279
; TYPE: DNA
```


us-09-656-668-198.rni

Wed Nov 7 09:21:20 2001

Query Match
Best Local Similarity 62.6%; Pred. No. 0.011;
Matches 77; Conservative 0; Mismatches 44; Indels 2; Gaps 1;
US-08-481-968A-4

Query Match 6.4%; Score 38.6; DB 4; Length 11288;

Best Local Similarity 62.6%; Pred. No. 0.011;
Matches 77; Conservative 0; Mismatches 44; Indels 2; Gaps 1;

QY 25 ccagtgatattgcaattcctaagacgtgtttgattgtcacacccctgggtgggaaca 84

DB 4309 CCAGGGACATTGGAAAGCTCGAGACATTGTTGTTTACAGCTTGGGGAGGA--G 4252

QY 85 tgcactggcatcatgcatagagggcgagtaatgctgtaaacatctttcaacycacag 144

DB 4251 TCCTGTGCGCATTTAGTGGGTAGATCCCCAGATGGTCTGAACACCCCTACACTTACAG 4192

QY 145 gac 147

DB 4191 AAC 4189

RESULT 9

US-08-910-733-17/c
; Sequence 17, Application US/08910733
; Patent No. 5837495

GENERAL INFORMATION:

APPLICANT: COLOTTA, Francesco

APPLICANT: MUZIO, Marta

APPLICANT: MANTOVANI, Alberto

TITLE OF INVENTION: INTERLEUKIN-1 ANTAGONIST, DNA ENCODING SAME,

TITLE OF INVENTION: AND ANTIBODIES THERETO

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK

STREET: 419 Seventh Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/910,733

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/476,860

FILING DATE: 07-JUN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: IT MI 94 A 002097

FILING DATE: 13-OCT-1994

ATTORNEY/AGENT INFORMATION:

NAME: YUN, Allen C.

REGISTRATION NUMBER: 37,971

REFERENCE/DOCKET NUMBER: COLOTTA-1A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:

LENGTH: 3176 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US 08-910-733-17

Query Match 6.0%; Score 36.6; DB 2; Length 3176;

Best Local Similarity 65.1%; Pred. No. 0.026;

Matches 54; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 387 agaatacacaaccatcctggtggcctgggactgtcctagtttagcattgaaagtcca 446

DB 2648 AGAGTGACCAACTACCTTGTGTTGCCAGGACTTCTTGACTTTATTACTGAAAGTTCTA 2589

QY 447 ggttcaggaagaaagccctcaggcc 469

DB 2588 CACCTTGCCAAACCCCTCAGTCC 2566

RESULT 10

US-08-910-884-17/c

; Sequence 17, Application US/08910884

; Patent No. 5981713

GENERAL INFORMATION:

APPLICANT: COLOTTA, Francesco

APPLICANT: MUZIO, Marta

APPLICANT: MANTOVANI, Alberto

TITLE OF INVENTION: INTERLEUKIN-1 ANTAGONIST, DNA ENCODING SAME,

TITLE OF INVENTION: AND ANTIBODIES THERETO

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK

STREET: 419 Seventh Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/910,884

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/476,860

FILING DATE: 07-JUN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: IT MI 94 A 002097

FILING DATE: 13-OCT-1994

ATTORNEY/AGENT INFORMATION:

NAME: YUN, Allen C.

REGISTRATION NUMBER: 37,971

REFERENCE/DOCKET NUMBER: COLOTTA-1B

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:

LENGTH: 3176 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-08-910-884-17

Query Match 6.0%; Score 36.6; DB 2; Length 3176;

Best Local Similarity 65.1%; Pred. No. 0.026;

Matches 54; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 387 agaatacacaaccatcctggtggcctgggactgtcctagtttagcattgaaagtcca 446

DB 2648 AGAGTGACCAACTACCTTGTGTTGCCAGGACTTCTTGACTTTATTACTGAAAGTTCTA 2589

QY 447 ggttcaggaagaaagccctcaggcc 469

APPLICANT: Nilsson, Jeanette
 APPLICANT: Tornell, Jan
 TITLE OF INVENTION: Genomic DNA Sequences
 TITLE OF INVENTION: Encoding Human BSSL/CEL
 NUMBER OF SEQUENCES: 58
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: White & Case
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: United States
 ZIP: 10036-2787
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/442,806
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/068,945
 FILING DATE: 27-MAY-1993
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: SE 9201809-2
 FILING DATE: 11-JUN-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: SE 9201826-6
 FILING DATE: 12-JUN-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: SE 9202088-2
 FILING DATE: 03-JUL-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: SE 9300902-5
 FILING DATE: 19-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Sterner, Richard J.
 REGISTRATION NUMBER: 35,372
 REFERENCE/DOCKET NUMBER: 1103326-052
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212)819-8783
 TELEFAX: (212)354-8113
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 11531 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 TISSUE TYPE: Mammary gland
 FEATURE:
 NAME/KEY: CDS
 LOCATION: join(1653..1727, 4071..4221, 4307..4429, 4707
 LOCATION: ..4904, 6193..6323, 6501..6608, 6751..6868, 8335
 LOCATION: ..8521, 8719..8922, 10124..10321, 10650..11394)
 FEATURE:
 NAME/KEY: mat_peptide
 LOCATION: join(1722..1727, 4071..4221, 4307..4429, 4707
 LOCATION: ..4904, 6193..6323, 6501..6608, 6751..6868, 8335
 LOCATION: ..8521, 8719..8922, 10124..10321, 10650..11391)
 OTHER INFORMATION: /EC_number= 3.1.1.1
 OTHER INFORMATION: /product= "Bile Salt-Stimulated Lipase"
 FEATURE:
 NAME/KEY: 5'UTR
 LOCATION: 1..1640
 FEATURE:
 NAME/KEY: TATA_signal
 LOCATION: 1611..1617
 FEATURE:

NAME/KEY: exon
 LOCATION: 1641..1727
 FEATURE:
 NAME/KEY: exon
 LOCATION: 4071..4221
 FEATURE:
 NAME/KEY: exon
 LOCATION: 4307..4429
 FEATURE:
 NAME/KEY: exon
 LOCATION: 4707..4904
 FEATURE:
 NAME/KEY: exon
 LOCATION: 6193..6323
 FEATURE:
 NAME/KEY: exon
 LOCATION: 6501..6608
 FEATURE:
 NAME/KEY: exon
 LOCATION: 6751..6868
 FEATURE:
 NAME/KEY: exon
 LOCATION: 8335..8521
 FEATURE:
 NAME/KEY: exon
 LOCATION: 8719..8922
 FEATURE:
 NAME/KEY: exon
 LOCATION: 10124..10321
 FEATURE:
 NAME/KEY: exon
 LOCATION: 10650..11490
 FEATURE:
 NAME/KEY: 3'UTR
 LOCATION: 11491..11531
 US-08-442-806-1

Query Match 5.9%; Score 35.6; DB 1; Length 11531;
 Best Local Similarity 61.4%; Pred. No. 0.12;
 Matches 108; Conservative 0; Mismatches 64; Indels 4; Gaps 3;

QY 14 ttacccccatccagtgatatttgcaatttcctaaagacgtgttttattgtacacactg 73
 Db 8055 TGACTCTCAGCCTGGGGACGTGTGGCAGGGCTGGAGACAT-CTGTGATTGTACAGCTG 8113
 QY 74 ggtgggaacatgctactggcatctaatatgcataagggaggaagtaagtcgtaaacatctt 133
 Db 8114 GGGAGGG--GGTGTCTCTGGCACCTCTGTGGTGTGGGCGGGGATGCT-CTAAACATCCT 8170
 QY 134 tcaacgcacagggacagaccccccaaaaaagagaattatctagcccaaatgtccata 189
 Db 8171 ACAGGGCAGAGGATGCCCTGTGATGTGCAGAAATCAACCTGCCCCCAAGTGTCATA 8226

RESULT 13
 US-09-429-323-3
 ; Sequence 3, Application US/09429323A
 ; Patent No. 6140126
 ; Patent No. 6140126 6140123
 ; GENERAL INFORMATION:
 ; APPLICANT: C. Frank Bennett
 ; APPLICANT: Lex M. Cowser
 ; TITLE OF INVENTION: ANTISENSE MODULATION OF Y-BOX BINDING PROTEIN 1 EXPRESSION
 ; FILE REFERENCE: RTS-0092
 ; CURRENT APPLICATION NUMBER: US/09/429,323A
 ; CURRENT FILING DATE: 1999-10-26
 ; NUMBER OF SEQ ID NOS: 89
 ; SEQ ID NO 3
 ; LENGTH: 1481
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:

us-09-656-668-198.rni

Wed Nov 7 09:21:20 2001

```

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..1109
; US-08-474-379C-31

Query Match      5.8%; Score 35.4; DB 2; Length 3073;
Best Local Similarity 47.9%; Pred. No. 0.064;
Matches 102; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 70 cctgggtggggaacatgctactgcatcctaataatgcatagaggcagtaagtgcctaaaca 129
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 280 CCGCGCGGGGACAAAGAGGTCTATCGCAACGACGAGGTTTGGGAACAGTAAATGGTTCATG 339
QY 130 tctttcaacgcacagacagagcccccacacaaagagaattatctagcccccaaatgtccata 189
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 340 TAAGCAACGGATATGTTTTCATCAACAGGAATGACACCCAGGAGATGTATTGTACACC 399
QY 190 acactgctgttgagaaacctaccgcaggtatcttactgggcttcattaggttaagcttgct 249
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 400 AGACTGCCATAAAGAGAAATACCCAGGAGTACCTTCGCAGTGTAGGAGATGAGAGAGA 459
QY 250 ttgtctggctctctgtagatatataaaataaag 282
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 460 CTGTGGAGTTTGATGTTGTTGAAGGAGAAAAGG 492
```

Search completed: November 5, 2001, 22:58:05
Job time: 2485 sec